Requested Patent:

WO9219743

Title:

HCV GENOMIC SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS

Abstracted Patent:

WO9219743

Publication Date:

1992-11-12

inventor(s):

BEALL EILEEN (US); CHA TAI-AN (US); IRVINE BRUCE (US); KOLBERG JANICE (US); URDEA MICHAEL S (US)

Applicant(s):

CHIRON CORP (US)

Application Number:

WO1992US04036 19920508

Priority Number(s):

US19910697326 19910508

IPC Classification:

A61K39/29; C07K13/00; C12N15/40; C12N15/51; C12Q1/68; C12Q1/70; - G01N33/576

Equivalents:

AU2155892, BG101876, BG62142, CA2108466, CZ9601210, EP0585398, HU69609, PT100472

ABSTRACT:

The present application features nucleic acid, peptide and antibody compositions relating to genotypes of hepatitis C virus and methods of using such compositions for diagnostic and therapeutic purposes.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZAN



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) Internati nal Patent Classification ⁵: C12N 15/51, C12Q 1/68 C12N 15/40, C12Q 1/70 A61K 39/29, C07K 13/00 G01N 33/576

(11) International Publication Number:

WO 92/19743

(43) International Publication Date:

12 November 1992 (12.11.92)

(21) International Application Number:

PCT/US92/04036

A2

(22) International Filing Date:

8 May 1992 (08.05.92)

(30) Priority data:

697,326

8 May 1991 (08.05.91)

US

(71) Applicant: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).

(72) Inventors: CHA, Tai-An; 964 Springview Circle, San Ramon, CA 94583 (US). BEALL, Eileen; 1150 Lincoln Avenue, #5, Walnut Creek, CA 94596 (US). IRVINE, Bruce; 3401 El Monte Drive, Concord, CA 94519 (US). KOLBERG, Janice; 131 Scots Valley, Hercules, CA 94547 (US). URDEA, Michael, S.; 100 Bunce Meadow Road, Alamo, CA 94501 (US).

(74) Agent: JANIUK, Anthony, J.; Wolf, Greenfield & Sacks, 600 Atlantic Avenue, Boston, MA 02210 (US).

(81) Designated States: AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GA (OAPI patent), GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU (European patent), MC (European patent), MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL (European patent), NO, PI, RO, RU, SD, SE (European patent), SN (OAPI patent), TD (OAPI patent), TG (OAPI patent).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: HCV GENOMIC SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS

(57) Abstract

The present application features nucleic acid, peptide and antibody compositions relating to genotypes of hepatitis C virus and methods of using such compositions for diagnostic and therapeutic purposes.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	es Fl	Spain Finland	MG Ml.	Madagascar Mali
AU	Australia	FR	1-ranco	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium		United Kingdom	MW	Malawi
BF	Burkina Faso	GB		NL	Netherlands .
BG	Bulgaria	GN	Guinca	NO	Norway
BJ	Benin	GR	Greece	PL	Poland
BR	Rewil	HU	Hungary	RO	Romania
CA	Canada	ΙT	ltily	RU	Russian Federation
CF	Central African Republic	JP	Japan		Sudan
CG	Congo	KP	Democratic People's Republic	SD	
CH	Switzerland		of Korea	SE	Sweden
CI.	Cole d'Ivoire	KR	Republic of Korea	SN	Senegal
		LI	Liechtenstein	SU	Soviet Union
CM	Camerosa	ŁK	Sri Lanka	TD	Chud
cs	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Cicemany		Munaco	US	United States of America
DK	Denmark	MC	Minimo		

WO 92/19743 PCT/US92/04036

1

HCV GENOMIC SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS

This application is a continuation-in-part of U.S. Serial No. 07/697,326 entitled "Polynucleotide Probes Useful for Screening for Hepatitis C Virus, filed May 8, 1991.

Technical Field

The invention relates to compositions and methods for the detection and treatment of hepatitis C virus, (HCV) infection, formerly referred to as blood-borne non-A, non-B hepatitis virus (NANBV) infection. More specifically, embodiments of the present invention feature compositions and methods for the detection of HCV, and for the development of vaccines for the prophylactic treatment of infections of HCV, and development of antibody products for conveying passive immunity to HCV.

20

25

Background of the Invention

The prototype isolate of HCV was characterized in U.S. Patent Application Serial No. 122,714 (See also EPO Publication No. 318,216). As used herein, the term "HCV" includes new isolates of the same viral species. The term "HCV-1" referred to in U.S. Patent Application S rial No. 122,714.

10

15

20

25

HCV is a transmissible disease distinguishable from other forms of viral-associated liver diseases, including that caused by the known hepatitis viruses, i.e., hepatitis A virus (HAV), hepatitis B virus (HBV), and delta hepatitis virus (HDV), as well as the hepatitis induced by cytomegalovirus (CMV) or Epstein-Barr virus (EBV). HCV was first identified in individuals who had received blood transfusions.

The demand for sensitive, specific methods for screening and identifying carriers of HCV and HCV contaminated blood or blood products is significant. Post-transfusion hepatitis (PTH) occurs in approximately 10% of transfused patients, and HCV accounts for up to 90% of these cases. The disease frequently progresses to chronic liver damage (25-55%).

Patient care as well as the prevention of transmission of HCV by blood and blood products or by close personal contact require reliable screening, diagnostic and prognostic tools to detect nucleic acids, antigens and antibodies related to HCV.

Information in this application suggests the HCV has several genotypes. That is, the genetic information of the HCV virus may not be totally identical for all HCV, but encompasses groups with differing genetic information.

Genetic information is stored in thread-like molecules of DNA and RNA. DNA consists of covalently

linked chains of deoxyribonucleotides and RNA consists of covalently linked chains of ribonucleotides. Each nucleotide is characterized by one of four bases: adenine (A), guanine (G), thymine (T), and cytosine (C). The bases are complementary in the sense that, due to the orientation of functional groups, certain base pairs attract and bond to each other through hydrogen bonding and m-stacking interactions. Adenine in one strand of DNA pairs with thymine in an opposing complementary strand. Guanine in one strand 10 of DNA pairs with cytosine in an opposing complementary strand. In RNA, the thymine base is replaced by uracil (U) which pairs with adenine in an opposing complementary strand. The genetic code of living organism is carried in the sequence of base pairs. 15 Living cells interpret, transcribe and translate the information of nucleic acid to make proteins and peptides.

The HCV genome is comprised of a single positive
strand of RNA. The HCV genome possesses a continuous,
translational open reading frame (ORF) that encodes a
polyprotein of about 3,000 amino acids. In the ORF,
the structural protein(s) appear to be encoded in
approximately the first quarter of the N-terminus
region, with the majority of the polyprotein
responsible for non-structural proteins.

10

15

20

25

The HCV polyprotein comprises, from the amino terminus to the carboxy terminus, the nucleocapsid protein (C), the envelope protein (E), and the non-structural proteins (NS) 1, 2 (b), 3, 4 (b), and 5.

HCV of differing genotypes may encode for proteins which present an altered response to host immune systems. HCV of differing genotypes may be difficult to detect by immuno diagnostic techniques and nucleic acid probe techniques which are not specifically directed to such genotype.

Definitions for selected terms used in the application are set forth below to facilitate an understanding of the invention. The term "corresponding" means homologous to or complementary to a particular sequence of nucleic acid. As between nucleic acids and peptides, corresponding refers to amino acids of a peptide in an order derived from the sequence of a nucleic acid or its complement.

The term "non-naturally occurring nucleic acid" refers to a portion of genomic nucleic acid, cDNA, semisynthetic nucleic acid, or synthetic origin nucleic acid which, by virtue of its origin or manipulation:

(1) is not associated with all of a nucleic acid with which it is associated in nature, (2) is linked to a nucleic acid or other chemical agent other than that to

WO 92/19743 PCT/US92/04036

- 5 -

which it is linked in nature, or (3) does not occur in nature.

Similarly the term, "a non-naturally occurring peptide" refers to a portion of a large naturally occurring peptide or protein, or semi-synthetic or synthetic peptide, which by virtue of its origin or manipulation (1) is not associated with all of a peptide with which it is associated in nature, (2) is linked to peptides, functional groups or chemical agents other than that to which it is linked in nature, or (3) does not occur in nature.

5

10

15

The term "primer" refers to a nucleic acid which is capable of initiating the synthesis of a larger nucleic acid when placed under appropriate conditions. The primer will be completely or substantially complementary to a region of the nucleic acid to be copied. Thus, under conditions conducive to hybridization, the primer will anneal to a

complementary region of a larger nucleic acid. Upon
addition of suitable reactants, the primer is extended
by the polymerizing agent to form a copy of the larger
nucleic acid.

The term "binding pair" refers to any pair of molecules which exhibit mutual affinity or binding

25 capacity. For the purposes of the present application, the term "ligand" will refer to one molecule of the binding pair, and the term "antiligand" or "receptor"

15

20

or "target" will refer to the opposite molecule of the binding pair. For example, with respect to nucleic acids, a binding pair may comprise two complementary nucleic acids. One of the nucleic acids may be designated the ligand and the other strand is designated the antiligand receptor or target. The designation of ligand or antiligand is a matter of arbitrary convenience. Other binding pairs comprise, by way of example, antigens and antibodies, drugs and drug receptor sites and enzymes and enzyme substrates, to name a few.

The term "label" refers to a molecular moiety capable of detection including, by way of example, without limitation, radioactive isotopes, enzymes, luminescent agents, precipitating agents, and dyes.

The term "support" includes conventional supports such as filters and membranes as well as retrievable supports which can be substantially dispersed within a medium and removed or separated from the medium by immobilization, filtering, partitioning, or the like. The term "support means" refers to supports capable of being associated to nucleic acids, peptides or antibodies by binding partners, or covalent or noncovalent linkages.

A number of HCV strains and isolates have been identified. When compared with the sequence of the original isolate derived from the USA ("HCV-1"; see

Q.-L. Choo et al. (1989) Science 244:359-362, Q.-L. Choo et al. (1990) Brit. Med. Bull. 46:423-441, Q.-L. Choo et al., Proc. Natl. Acad. Sci. 88:2451-2455 (1991), and E.P.O. Patent Publication No. 318,216, cited supra), it was found that a Japanese isolate ("HCV J1") differed significantly in both nucleotide and polypeptide sequence within the NS3 and NS4 regions. This conclusion was later extended to the NS5 and envelope (E1/S and E2/NS1) regions (see K. Takeuchi 10 et al., J. Gen. Virol. (1990) 71:3027-3033, Y. Kubo, Nucl. Acids. Res. (1989) 17:10367-10372, and K. Takeuchi et al., Gene (1990) 91:287-291). The former group of isolates, originally identified in the United States, is termed "Genotype I" throughout the present 15 disclosure, while the latter group of isolates, initially identified in Japan, is termed "Genotype II" herein.

Brief Description of the Invention

The present invention features compositions of matter comprising nucleic acids and peptides corresponding to the HCV viral genome which define different genotypes. The present invention also features methods of using the compositions corresponding to sequences of the HCV viral genome which define different genotypes described herein.

10

15

20

25

A. Nucleic acid compositions

The nucleic acid of the present invention, corresponding to the HCV viral genome which define different genotypes, have utility as probes in nucleic acid hybridization assays, as primers for reactions involving the synthesis of nucleic acid, as binding partners for separating HCV viral nucleic acid from other constituents which may be present, and as anti-sense nucleic acid for preventing the transcription or translation of viral nucleic acid.

One embodiment of the present invention features a composition comprising a non-naturally occurring nucleic acid having a nucleic acid sequence of at least eight nucleotides corresponding to a non-HCV-1 nucleotide sequence of the hepatitis C viral genome. Preferably, the nucleotide sequence is selected from a sequence present in at least one region consisting of the NS5 region, envelope 1 region, 5'UT region, and the core region.

Preferably, with respect to sequences which correspond to the NS5 region, the sequence is selected from a sequence within a sequence numbered 2-22. The sequence numbered 1 corresponds to HCV-1. Sequences numbered 1-22 are defined in the Sequence Listing of the application.

Preferably, with respect to sequences corresponding to the envelope 1 region, the sequence is

selected from a sequence within sequences numbered 24-32. Sequence No. 23 corresponds to HCV-1. Sequences numbered 23-32 are set forth in the Sequence Listing of the application.

5

10

15

Preferably, with respect to the sequences which correspond to the 5'UT regions, the sequence is selected from a sequence within sequences numbered 34-51. Sequence No. 33 corresponds to HCV-1. Sequence No. 33-51 are set forth in the Sequence Listing of this application.

Preferably, with respect to the sequences which correspond to the core region, the sequence is selected from a sequence within the sequences numbered 53-66. Sequence No. 52 corresponds to HCV-1. Sequences 52-66 are set forth in the Sequence Listing of this application.

The compositions of the present invention form hybridization products with nucleic acid corresponding to different genotypes of HCV.

HCV has at least five genotypes, which will be referred to in this application by the designations GI-GV. The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. The second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV,

10

15

20

25

is exemplified by sequences numbered 20-22, and 29-31 and 48-49. The fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51.

One embodiment of the present invention features compositions comprising a nucleic acid having a sequence corresponding to one or more sequences which exemplify a genotype of HCV.

B. Method of forming a Hybridization Product

Embodiments of the present invention also feature a method of forming a hybridization product with nucleic acid having a sequence corresponding to HCV nucleic acid. One method comprises the steps of placing a non-naturally occurring nucleic acid having a non-HCV-I sequence corresponding to HCV nucleic acid under conditions in which hybridization may occur. The non-naturally occurring nucleic acid is capable of forming a hybridization product with HCV nucleic acid, under hybridization conditions. The method further comprises the step of imposing hybridization conditions to form a hybridization product in the presence of nucleic acid corresponding to a region of the HCV genome.

The formation of a hybridization product has utility for detecting the presence of one or more genotypes of HCV. Preferably, the non-naturally occurring nucleic acid forms a hybridization product

with nucleic acid of HCV in one or more regions comprising the NS5 region, envelope 1 region, 5'UT region and the core region. To detect the hybridization product, it is useful to associate the non-naturally occurring nucleic acid with a label. The formation of the hybridization product is detected by separating the hybridization product from labeled non-naturally occurring nucleic acid, which has not formed a hybridization product.

5

10

15

20

25

The formation of a hybridization product has utility as a means of separating one or more genotypes of HCV nucleic acid from other constituents potentially present. For such applications, it is useful to associate the non-naturally occurring nucleic acid with a support for separating the resultant hybridization product from the the other constituents.

Nucleic acid "sandwich assays" employ one nucleic acid associated with a label and a second nucleic acid associated with a support. An embodiment of the present invention features a sandwich assay comprising two nucleic acids, both have sequences which correspond to HCV nucleic acids; however, at least one non-naturally occurring nucleic acid has a sequence corresponding to non-HCV-1 HCV nucleic acid. At least one nucleic acid is capable of associating with a label, and the other is capable of associating with a support. The support associated non-naturally

10

15

20

25

occurring nucleic acid is used to separate the hybridization products which include an HCV nucleic acid and the non-naturally occurring nucleic acid having a non-HCV-1 sequence.

One embodiment of the present invention features a method of detecting one or more genotypes of HCV. method comprises the steps of placing a non-naturally occurring nucleic acid under conditions which hybridization may occur. The non-naturally occurring nucleic acid is capable of forming a hybridization product with nucleic acid from one or more genotypes of HCV. The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV, is exemplified sequences numbered 20-22 and 29-31. The fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51.

The hybridization product of HCV nucleic acid with a non-naturally occurring nucleic acid having non-HCV-1 sequence corresponding to sequences within the HCV genome has utility for priming a reaction for the synthesis of nucleic acid.

The hybridization product of HCV nucleic acid with a non-naturally occurring nucleic acid having a

10

sequence corresponding to a particular genotype of HCV has utility for priming a reaction for the synthesis of nucleic acid of such genotype. In one embodiment, the synthesized nucleic acid is indicative of the presence of one or more genotypes of HCV.

The synthesis of nucleic acid may also facilitate cloning of the nucleic acid into expression vectors which synthesize viral proteins.

Embodiments of the present methods have utility as anti-sense agents for preventing the transcription or translation of viral nucleic acid. The formation of a hybridization product of a non-naturally occurring nucleic acid having sequences which correspond to a particular genotype of HCV genomic sequencing with HCV 15 nucleic acid may block translation or transcription of such genotype. Therapeutic agents can be engineered to include all five genotypes for inclusivity.

C. Peptide and antibody composition

A further embodiment of the present invention 20 features a composition of matter comprising a non-naturally occurring peptide of three or more amino acids corresponding to a nucleic acid having a non-HCV-1 sequence. Preferably, the non-HCV-1 sequence corresponds with a sequence within one or more regions 25 consisting of the NS5 region, the envelope 1 region, the 5'UT region, and the core region.

15

Preferably, with respect to peptides corresponding to a nucleic acid having a non-HCV-1 sequence of the NS5 region, the sequence is within sequences numbered 2-22. The sequence numbered 1 corresponds to HCV-1. Sequences numbered 1-22 are set forth in the Sequence Listing.

Preferably, with respect to peptides corresponding to a nucleic acid having a non-HCV-1 sequence of the envelope 1 region, the sequence is within sequences numbered 24-32. The sequence numbered 23 corresponds to HCV-1. Sequences numbered 23-32 are set forth in the Sequence Listing.

Preferably, with respect to peptides corresponding to a nucleic acid having a non-HCV-1 sequence directed to the core region, the sequence is within sequences numbered 53-66. Sequence numbered 52 corresponds to HCV-1. Sequences numbered 52-66 are set forth in the Sequence Listing.

features peptide compositions corresponding to nucleic acid sequences of a genotype of HCV. The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. The second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV, is exemplified

sequences numbered 20-22, 29-31, 48 and 49. The fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51.

5

10

15

20

25

The non-naturally occurring peptides of the present invention are useful as a component of a vaccine. The sequence information of the present invention permits the design of vaccines which are inclusive for all or some of the different genotypes of HCV. Directing a vaccine to a particular genotype allows prophylactic treatment to be tailored to maximize the protection to those agents likely to be encountered. Directing a vaccine to more than one genotype allows the vaccine to be more inclusive.

The peptide compositions are also useful for the development of specific antibodies to the HCV proteins. One embodiment of the present invention features as a composition of matter, an antibody to peptides corresponding to a non-HCV-1 sequence of the HCV genome. Preferably, the non-HCV-1 sequence is selected from the sequence within a region consisting of the NS5 region, the envelope 1 region, and the core region. There are no peptides associated with the untranslated 5'UT region.

Preferably, with respect to antibodies directed to peptides of the NS5 region, the peptide corresponds to a sequence within sequences numbered 2-22. Preferably, with respect to antibodies directed to a peptide

10

15

20

25

corresponding to the envelope 1 region, the peptide corresponds to a sequence within sequences numbered 24-32. Preferably, with respect to the antibodies directed to peptides corresponding to the core region, the peptide corresponds to a sequence within sequences numbered 53-66.

Antibodies directed to peptides which reflect a particular genotype have utility for the detection of such genotypes of HCV and therapeutic agents.

One embodiment of the present invention features an antibody directed to a peptide corresponding to nucleic acid having sequences of a particular genotype. The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. The second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV, is exemplified sequences numbered 20-22, 29-31, 48 and 49. The fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51.

Individuals skilled in the art will readily recognize that the compositions of the present invention can be packaged with instructions for use in the form of a kit for performing nucleic acid hybridizations or immunochemical reactions.

WO 92/19743 PCT/US92/04036

- 17 -

The present invention is further described in the following figures which illustrate sequences demonstrating genotypes of HCV. The sequences are designated by numerals 1-145, which numerals and sequences are consistent with the numerals and sequences set forth in the Sequence Listing. Sequences 146 and 147 facilitate the discussion of an assay which numerals and sequences are consistent with the numerals and sequences set forth in the Sequence Listing.

10

25

5

Brief Description of the Figures and Sequence Listing

Figure 1 depicts schematically the genetic organization of HCV;

Figure 2 sets forth nucleic acid sequences

numbered 1-22 which sequences are derived from the NS5
region of the HCV viral genome;

Figure 3 sets forth nucleic acid sequences numbered 23-32 which sequences are derived from the envelope 1 region of the HCV viral genome;

20 Figure 4 sets forth nucleic acid sequences numbered 33-51 which sequences are derived from the 5'UT region of the HCV viral genome; and,

Figure 5 sets forth nucleic acid sequences numbered 52-66 which sequences are derived from the core region of the HCV viral genome.

The Sequence Listing sets forth the sequences of sequences numbered 1-147.

10

15

Detailed Description of the Invention

The present invention will be described in detail as as nucleic acid having sequences corresponding to the HCV genome and related peptides and binding partners, for diagnostic and therapeutic applications.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Maniatis, Fitsch & Sambrook, Molecular Cloning; A Laboratory Manual (1982); DNA Cloning, Volumes I and II (D.N Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed, 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 1984); the series, Methods in Enzymology (Academic Press, Inc.), particularly Vol. 154 and Vol. 155 (Wu and Grossman, eds.).

The cDNA libraries are derived from nucleic acid sequences present in the plasma of an HCV-infected chimpanzee. The construction of one of these libraries, the "c" library (ATCC No. 40394), is described in PCT Pub. No. WO90/14436. The sequences of the library relevant to the present invention are set forth herein as sequence numbers 1, 23, 33 and 52.

Nucleic acids isolated or synthesized in accordance with features of the present invention are

useful, by way of example without limitation as probes, primers, anti-sense genes and for developing expression systems for the synthesis of peptides corresponding to such sequences.

The nucleic acid sequences described define genotypes of HCV with respect to four regions of the viral genome. Figure 1 depicts schematically the organization of HCV. The four regions of particular interest are the NS5 region, the envelope 1 region, the 5'UT region and the core region.

5

10

15

20

25

The sequences set forth in the present application as sequences numbered 1-22 suggest at least five genotypes in the NS5 region. Sequences numbered 1-22 are depicted in Figure 2 as well as the Sequence Listing. Each sequence numbered 1-22 is derived from nucleic acid having 340 nucleotides from the NS5 region.

The five genotypes are defined by groupings of the sequences defined by sequence numbered 1-22. For convenience, in the present application, the different genotypes will be assigned roman numerals and the letter "G".

The first genotype (GI) is exemplified by sequences within sequences numbered 1-6. A second genotype (GII) is exemplified by sequences within sequences numbered 7-12. A third genotype (GIII) is exemplified by the sequences within sequences numbered 13-17. A fourth genotype (GIV) is exemplified by

10

15

sequences within sequences numbered 20-22. A fifth genotype (GV) is exemplified by sequences within sequences numbered 18 and 19.

The sequences set forth in the present application as sequences numbered 23-32 suggest at least four genotypes in the envelope 1 region of HCV. numbered 23-32 are depicted in Figure 3 as well as in the Sequence Listing. Each sequence numbered 23-32 is derived from nucleic acid having 100 nucleotides from the envelope 1 region.

A first envelope 1 genotype group (GI) is exemplified by the sequences within the sequences numbered 23-25. A second envelope 1 genotype (GII) region is exemplified by sequences within sequences numbered 26-28. A third envelope 1 genotype (GIII) is exemplified by the sequences within sequences numbered 32. A fourth envelope 1 genotype (GIV) is exemplified by the sequences within sequence numbered 29-31.

The sequences set forth in the present application as sequences numbered 33-51 suggest at least three 20 genotypes in the 5'UT region of HCV. Sequences numbered 33-51 are depicted in Figure 4 as well as in the Sequence Listing. Each sequence numbered 33-51 is derived from the nucleic acid having 252 nucleotides from the 5'UT region, although sequences 50 and 51 are 25 somewhat shorter at approximately 180 nucleotides.

10

15

20

25

The first 5'UT genotype (GI) is exemplified by the sequences within sequences numbered 33-38. A second 5'UT genotype (GII) is exemplified by the sequences within sequences numbered 39-45. A third 5'UT genotype (GIII) is exemplified by the sequences within sequences numbered 46-47. A fourth 5'UT genotype (GIV) is exemplified by sequences within sequences humbered 48 and 49. A fifth 5'UT genotype (GV) is exemplified by sequences within sequences numbered 50 and 51.

The sequences numbered 48-62 suggest at least three genotypes in the core region of HCV. The sequences numbered 52-66 are depicted in Figure 5 as well as in the Sequence Listing.

The first core region genotype (GI) is exemplified by the sequences within sequences numbered 52-57. The second core region genotype (GII) is exemplified by sequences within sequences numbered 58-64. The third core region genotype (GIII) is exemplified by sequences within sequences numbered 65 and 66. Sequences numbered 52-65 are comprised of 549 nucleotides. Sequence numbered 66 is comprised of 510 nucleotides.

The various genotypes described with respect to each region are consistent. That is, HCV having features of the first genotype with respect to the NS5 region will substantially conform to features of the first genotype of the envelope 1 region, the 5'UT region and the core region.

Nucleic acid isolated or synthesized in accordance with the sequences set forth in sequence numbers 1-66 are useful as probes, primers, capture ligands and anti-sense agents. As probes, primers, capture ligands and anti-sense agents, the nucleic acid wil normally comprise approximately eight or more nucleotides for specificity as well as the ability to form stable hybridization products.

10 Probes

15

20

A nucleic acid isolated or synthesized in accordance with a sequence defining a particular genotype of a region of the HCV genome can be used as a probe to detect such genotype or used in combination with other nucleic acid probes to detect substantially all genotypes of HCV.

With the sequence information set forth in the present application, sequences of eight or more nucleotides are identified which provide the desired inclusivity and exclusivity with respect to various genotypes within HCV, and extraneous nucleic acid sequences likely to be encountered during hybridization conditions.

Individuals skilled in the art will readily
recognize that the nucleic acid sequences, for use as
probes, can be provided with a label to facilitate
detection of a hybridization product.

Capture Ligand

For use as a capture ligand, the nucleic acid selected in the manner described above with respect to probes, can be readily associated with supports. The manner in which nucleic acid is associated with supports is well known. Nucleic acid having sequences corresponding to a sequence within sequences numbered 1-66 have utility to separate viral nucleic acid of one genotype from the nucleic acid of HCV of a different genotype. Nucleic acid isolated or synthesized in accordance with sequences within sequences numbered 1-66, used in combinations, have utility to capture substantially all nucleic acid of all HCV genotypes.

15 Primers

10

20

25

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility as primers for the amplification of HCV sequences. With respect to polymerase chain reaction (PCR) techniques, nucleic acid sequences of eight or more nucleotides corresponding to one or more sequences of sequences numbered 1-66 have utility in conjunction with suitable enzymes and reagents to create copies of the viral nucleic acid. A plurality of primers having different sequences corresponding to more than one genotype can be used to create copies of viral nucleic acid for such genotypes.

10

15

20

The copies can be used in diagnostic assays to detect HCV virus. The copies can also be incorporated into cloning and expression vectors to generate polypeptides corresponding to the nucleic acid synthesized by PCR, as will be described in greater detail below.

Anti-sense

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility as anti-sense genes to prevent the expression of HCV.

Nucleic acid corresponding to a genotype of HCV is loaded into a suitable carrier such as a liposome for introduction into a cell infected with HCV. A nucleic acid having eight or more nucleotides is capable of binding to viral nucleic acid or viral messenger RNA. Preferably, the anti-sense nucleic acid is comprised of 30 or more nucleotides to provide necessary stability of a hybridization product of viral nucleic acid or viral messenger RNA. Methods for loading anti-sense nucleic acid is known in the art as exemplified by U.S. Patent 4,241,046 issued December 23, 1980 to Papahadjopoulos et al.

25 Peptide Synthesis

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility to

generate peptides. The sequences exemplified by sequences numbered 1-32 and 52-66 can be cloned into suitable vectors or used to isolate nucleic acid. isolated nucleic acid is combined with suitable DNA linkers and cloned into a suitable vector. can be used to transform a suitable host organism such as E. coli and the peptide encoded by the sequences isolated.

Molecular cloning techniques are described in the text Molecular Cloning: A Laboratory Manual, Maniatis 10 et al., Coldspring Harbor Laboratory (1982).

The isolated peptide has utility as an antigenic substance for the development of vaccines and antibodies directed to the particular genotype of HCV.

15

20

Vaccines and Antibodies

The peptide materials of the present invention have utility for the development of antibodies and vaccines.

The availability of cDNA sequences, or nucleotide sequences derived therefrom (including segments and modifications of the sequence), permits the construction of expression vectors encoding antigenically active regions of the peptide encoded in either strand. The antigenically active regions may be 25 derived from the NS5 region, envelope 1 regions, and the core region.

15

20

25

Fragments encoding the desired peptides are derived from the cDNA clones using conventional restriction digestion or by synthetic methods, and are ligated into vectors which may, for example, contain portions of fusion sequences such as beta galactosidase or superoxide dismutase (SOD), preferably SOD. Methods and vectors which are useful for the production of polypeptides which contain fusion sequences of SOD are described in European Patent Office Publication number 0196056, published October 1, 1986.

Any desired portion of the HCV cDNA containing an open reading frame, in either sense strand, can be obtained as a recombinant peptide, such as a mature or fusion protein; alternatively, a peptide encoded in the cDNA can be provided by chemical synthesis.

The DNA encoding the desired peptide, whether in fused or mature form, and whether or not containing a signal sequence to permit secretion, may be ligated into expression vectors suitable for any convenient host. Both eukaryotic and prokaryotic host systems are presently used in forming recombinant peptides. The peptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification may be by techniques known in the art, for example, differential extraction, salt fractionation, chromatography on ion exchange resins, affinity chromatography, centrifugation, and

the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins. Such peptides can be used as diagnostics, or those which give rise to neutralizing antibodies may be formulated into vaccines. Antibodies raised against these peptides can also be used as diagnostics, or for passive immunotherapy or for isolating and identifying HCV.

An antigenic region of a peptide is generally 10 relatively small--typically 8 to 10 amino acids or less in length. Fragments of as few as 5 amino acids may characterize an antigenic region. These segments may correspond to NS5 region, envelope 1 region, and the core region of the HCV genome. The 5'UT region is not known to be translated. Accordingly, using the cDNAs 15 of such regions, DNAs encoding short segments of HCV peptides corresponding to such regions can be expressed recombinantly either as fusion proteins, or as isolated In addition, short amino acid sequences can be conveniently obtained by chemical synthesis. 20 instances wherein the synthesized peptide is correctly configured so as to provide the correct epitope, but is too small to be immunogenic, the peptide may be linked to a suitable carrier.

25 A number of techniques for obtaining such linkage are known in the art, including the formation of disulfide linkages using N-succinimidyl-3-(2-

pyridylthio)propionate (SPDP) and succinimidyl 4-(N-maleimido-methyl)cyclohexane-l-carboxylate (SMCC) obtained from Pierce Company, Rockford, Illinois, (if the peptide lacks a sulfhydryl group, this can be provided by addition of a cysteine residue). reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an amide linkage through the epsilon-amino on a lysine, or other free amino group in the other. A variety of such disulfide/amide-forming agents are known. See, for 10 example, Immun Rev (1982) 62:185. Other bifunctional coupling agents form a thioether rather than a disulfide linkage. Many of these thio-ether-forming agents are commercially available and include reactive esters of 6-maleimidocaprioc acid, 2-bromoacetic acid, 15 2-iodoacetic acid, 4-N-maleimido-methyl)cyclohexane-lcarboxylic acid, and the like. The carboxyl groups can be activated by combining them with succinimide or 1-hydroxyl-2 nitro-4-sulfonic acid, sodium salt. Additional methods of coupling antigens employs the 20 rotavirus/"binding peptide" system described in EPO Pub. No. 259,149, the disclosure of which is incorporated herein by reference. The foregoing list is not meant to be exhaustive, and modifications of the named compounds can clearly be used. 25

Any carrier may be used which does not itself induce the production of antibodies harmful to the

10

host. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins; polysaccharides, such as latex functionalized Sepharose, agarose, cellulose, cellulose beads and the like; polymeric amino acids, such as polyglutamic acid, polylysine, and the like; amino acid copolymers; and inactive virus particles. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those skilled in the art.

Peptides comprising HCV amino acid sequences encoding at least one viral epitope derived from the NS5, envelope 1, and core region are useful immunological reagents. The 5'UT region is not known 15 to be translated. For example, peptides comprising such truncated sequences can be used as reagents in an immunoassay. These peptides also are candidate subunit antigens in compositions for antiserum production or vaccines. While the truncated sequences can be 20 produced by various known treatments of native viral protein, it is generally preferred to make synthetic or recombinant peptides comprising HCV sequence. Peptides comprising these truncated HCV sequences can be made up entirely of HCV sequences (one or more epitopes, either 25 contiguous or noncontiguous), or HCV sequences and heterologous sequences in a fusion protein. Useful

25

heterologous sequences include sequences that provide for secretion from a recombinant host, enhance the immunological reactivity of the HCV epitope(s), or facilitate the coupling of the polypeptide to an immunoassay support or a vaccine carrier. See, E.G., EPO Pub. No. 116,201; U.S. Pat. No. 4,722,840; EPO Pub. No. 259,149; U.S. Pat. No. 4,629,783.

The size of peptides comprising the truncated HCV sequences can vary widely, the minimum size being a sequence of sufficient size to provide an HCV epitope, 10 while the maximum size is not critical. For convenience, the maximum size usually is not substantially greater than that required to provide the desired HCV epitopes and function(s) of the heterologous sequence, if any. Typically, the 15 truncated HCV amino acid sequence will range from about 5 to about 100 amino acids in length. More typically, however, the HCV sequence will be a maximum of about 50 amino acids in length, preferably a maximum of about 30 amino acids. It is usually desirable to select HCV 20 sequences of at least about 10, 12 or 15 amino acids, up to a maximum of about 20 or 25 amino acids.

HCV amino acid sequences comprising epitopes can be identified in a number of ways. For example, the entire protein sequence corresponding to each of the NS5, envelope 1, and core regions can be screened by preparing a series of short peptides that together span the entire protein sequence of such regions. By starting with, for example, peptides of approximately 100 amino acids, it would be routine to test each peptide for the presence of epitope(s) showing a desired reactivity, and then testing progressively smaller and overlapping fragments from an identified peptides of 100 amino acids to map the epitope of interest. Screening such peptides in an immunoassay is within the skill of the art. It is also known to carry out a computer analysis of a protein sequence to identify potential epitopes, and then prepare peptides comprising the identified regions for screening.

10

The immunogenicity of the epitopes of HCV may also be enhanced by preparing them in mammalian or yeast 15 systems fused with or assembled with particle-forming proteins such as, for example, that associated with hepatitis B surface antigen. See, e.g., US 4,722,840. Constructs wherein the HCV epitope is linked directly to the particle-forming protein coding sequences 20 produce hybrids which are immunogenic with respect to the HCV epitope. In addition, all of the vectors prepared include epitopes specific to HBV, having various degrees of immunogenicity, such as, for example, the pre-S peptide. Thus, particles 25 constructed from particle forming protein which include HCV sequences are immunogenic with respect to HCV and HBV.

Hepatitis surface antigen (HBSAg) has been shown to be formed and assembled into particles in S. cerevisiae (P. Valenzuela et al. (1982)), as well as in, for example, mammalian cells (P. Valenzuela et al. 1984)). The formation of such particles has been shown to enhance the immunogenicity of the monomer subunit. The constructs may also include the immunodominant epitope of HBSAg, comprising the 55 amino acids of the presurface (pre-S) region. Neurath et al. (1984). Constructs of the pre-S-HBSAg particle expressible in 10 yeast are disclosed in EPO 174,444, published March 19, 1986; hybrids including heterologous viral sequences for yeast expression are disclosed in EPO 175,261, published March 26, 1966. These constructs may also be expressed in mammalian cells such as Chinese hamster 15 ovary (CHO) cells using an SV40-dihydrofolate reductase vector (Michelle et al. (1984)).

In addition, portions of the particle-forming protein coding sequence may be replaced with codons encoding an HCV epitope. In this replacement, regions which are not required to mediate the aggregation of the units to form immunogenic particles in yeast of mammals can be deleted, thus eliminating additional HBV antigenic sites from competition with the HCV epitope.

25

20

Vaccines

Vaccines may be prepared from one or more

immunogenic peptides derived from HCV. The observed homology between HCV and Flaviviruses provides information concerning the peptides which are likely to be most effective as vaccines, as well as the regions of the genome in which they are encoded.

Multivalent vaccines against HCV may be comprised of one or more epitopes from one or more proteins derived from the NS5, envelope 1, and core regions. In particular, vaccines are contemplated comprising one or more HCV proteins or subunit antigens derived from the NS5, envelope 1, and core regions. The 5'UT region is not known to be translated.

10

The preparation of vaccines which contain an immunogenic peptide as an active ingredient, is known to one skilled in the art. Typically, such vaccines 15 are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or 20 the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In 25 addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or

15

20

25

emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-theronyl-D- isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl- D-isoglutamine (CGP 11637, referred to as nor-MDP), N- acetylmuramyl-Lalanyl-D-isoglutaminyl-L-alanine-2-(1- 2-dipalmitoyl -sn-glycero-3-hydroxyphosphoryloxy)- ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against an immunogenic peptide containing an HCV antigenic sequence resulting from administration of this peptide in vaccines which are also comprised of the various adjuvants.

The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such

suppositories may be formed from mixtures containing the active ingredient in the range of 0/5% to 10%, preferably 1%-2%. Oral formulations include such normally employed excipients as, for example,

5 pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like.

The examples below are provided for illustrative purposes and are not intended to limit the scope of the present invention.

I. Detection of HCV RNA from Serum

RNA was extracted from serum using quanidinium salt, phenol and chloroform according to the

instructions of the kit manufacturer (RNAzol B kit, Cinna/Biotecx). Extracted RNA was precipitated with isopropanol and washed with ethanol. A total of 25 µl serum was processed for RNA isolation, and the purified RNA was resuspended in 5 µl diethyl

pyrocarbonate treated water for subsequent cDNA synthesis.

II. <u>cDNA Synthesis and Polymerase Chain Reaction (PCR)</u> Amplification

Table 1 lists the sequence and position (with reference to HCV1) of all the PCR primers and probes used in these examples. Letter designations for

nucleotides are consistent with 37 C.F.R. §§1.821—1.825. Thus, the letters A, C, G, T, and U are used in the ordinary sense of adenine, cytosine, quanine, thymine, and uracil. The letter M means A or C; R means A or G; W means A or T/U; S means C or G; Y means C or T/U; K means G or T/U; V means A or C or G, not T/U; H means A or C or T/U, not G; D means A or G or T/U, not C; B means C or G or T/U, not A; N means (A or C or G or T/U) or (unknown or other). Table 1 is set forth below:

Table 1

	Seq. No.	Table 1 Sequence (5'-3')	Nucleotide	Position
15 20	67 68 69 70 71 72 73	CAAACGTAACACCAACCGRCGCCCACAGG ACAGAYCCGCAKAGRTCCCCCACG GCAACCTCGAGGTAGACGTCAGCCTATCG GCAACCTCGTGGAAGGCGACAACCTATCG GTCACCAATGATTGCCCTAACTCGAGTAGTCACGAACGACGACTGCTCCAACTCAAG TGGACATGATCGCTGGWGCYCACTGGGG TGGAYATGGTGGYGGGGGCYCACTGGGG	11 CC 50 CC 50 TT 94	4-402 92-1169 99-538 99-538 48-977 48-973 375-1402 375-1402 308-1327
25	75 76 77 78 79 80	ATGATGAACTGGTCVCCYAC ACCTTVGCCCAGTTSCCCRCCATGGA AACCCACTCTATGYCCGGYCAT GAATCGCTGGGGTGACCG CCATGAATCACTCCCCTGTGAGGAACTA	1 2 1 3	453-1428 05-226 71-188 0-57

For cDNA synthesis and PCR amplification, a protocol developed by Perkin-Elmer/Cetus (GeneAmp® RNA PCR kit) was used. Both random hexamer and primers with specific complementary sequences to HCV were 5 employed to prime the reverse transcription (RT) reaction. All processes, except for adding and mixing reaction components, were performed in a thermal cycler (MJ Research, Inc.). The first strand cDNA synthesis reaction was inactivated at 99°C for 5 min, and then cooled at 50°C for 5 min before adding reaction 10 components for subsequent amplification. After an initial 5 cycles of 97°C for 1 min, 50°C for 2 min, and 72°C for 3 min, 30 cycles of 94°C for 1 min, 55°C for 2 min, and 72°C for 3 min followed, and then a final 7 min of elongation at 72°C. 15

For the genotyping analysis, sequences 67 and 68 were used as primers in the PCR reaction. These primers amplify a segment corresponding to the core and envelope regions. After amplification, the reaction products were separated on an agarose gel and then transferred to a nylon membrane. The immobilized reaction products were allowed to hybridize with a 32p-labelled nucleic acid corresponding to either Genotype I (core or envelope 1) or Genotype II (core or envelope 1). Nucleic acid corresponding to Genotype 1 comprised sequences numbered 69 (core), 71 (envelope), and 73 (envelope). Nucleic acid corresponding to

20

25

10

15

Genotype II comprised sequences numbered 70 (core), 72 (envelope), and 74 (envelope).

The Genotype I probes only hybridized to the product amplified from isolates which had Genotype I sequence. Similarly, Genotype II probes only hybridized to the product amplified from isolates which had Genotype II sequence.

In another experiment, PCR products were generated using sequences 79 and 80. The products were analyzed as described above except Sequence No. 73 was used to detect Genotype I, Sequence No. 74 was used to detect Genotype II, Sequence No. 77 (5'UT) was used to detect Genotype III, and Sequence No. 78 (5'UT) was used to detect Genotype IV. Each sequence hybridized in a genotype specific manner.

III. <u>Detection of HCV GI-GIV using a sandwich</u> hybridization assay for HCV RNA

An amplified solution phase nucleic acid sandwich
hybridization assay format is described in this
example. The assay format employs several nucleic acid
probes to effect capture and detection. A capture
probe nucleic acid is capable of associating a
complementary probe bound to a solid support and HCV
nucleic acid to effect capture. A detection probe
nucleic acid has a first segment (A) that binds to HCV
nucleic acid and a second segment (B) that hybridizes
to a second amplifier nucleic acid.

10

The amplifier nucleic acid has a first segment (B*) that hybridizes to segment (B) of the probe nucleic acid and also comprises fifteen iterations of a segment (C). Segment C of the amplifier nucleic acid is capable of hybridizing to three labeled nucleic acids.

Nucleic acid sequences which correspond to nucleotide sequences of the envelope 1 gene of Group I HCV isolates are set forth in sequences numbered 81-99. Table 2 sets forth the area of the HCV genome to which the nucleic acid sequences correspond and a preferred use of the sequences.

Table 2 Complement of Sequence No. Probe Type Nucleotide Numbers 15 879-911 Label 81 912-944 Label 82 945-977 Capture 83 978-1010 Label 84 20 1011-1043 Label 85 1044-1076 Label 86 1077-1109 Label 87 1110-1142 Capture 88 1143-1175 Label 89 25

- 40 -

Table 2 continued

	Probe Type	Sequence No.	Complement of Nucleotide Numbers
5	Label	90	1176-1208
	Label	91	1209-1241
•	Label	92	1242=1274
	Capture	93	1275-1307
10	Label	94	1308-1340
10	Label	95	1341–1373
	. Label	· 96	1374-1406
	Label	97	1407-1439
	Capture	98	1440-1472
15	Label	99	1473-1505

Nucleic acid sequences which correspond to nucleotide sequences of the envelope 1 gene of Group II HCV isolates are set forth in sequences 100-118. Table 3 sets forth the area of the HCV genome to which the nucleic acid corresponds and the preferred use of the sequences.

20

Table 3

	Probe Type	Sequence No.	Complement of Nucleotide Numbers
5	Label	100	879-911
	Label	101	912-944
	Capture	102	945-977
	Label	103	978-1010
10	Label	104	1011-1043
	Label	105	1044-1076
	Label	106	1077-1109
	Capture	107	1110-1142
	Label	108	1143-1175
15	Label	109	1176-1208
	Label	110	1209-1241
	Label	111	1242=1274
	Capture	112	1275-1307
	Label	113	1308-1340
20	Label	114	1341-1373
	Label	115	1374-1406
	Label	116	1407-1439
	Capture	117	1440-1472
	Label	118	1473-1505
25			

Nucleic acid sequences which correspond to nucleotide sequences in the C gene and the 5'UT region

are set forth in sequences 119-145. Table 4 identifies the sequence with a preferred use.

Table 4

5		
	Probe Type	Sequence No.
	Capture	119
•	Label	120
10	Label	121
•	Label	122
·	Capture	123
	Label	124
	Label	125
15	Label	126
	Capture	127
	Label	128
	Label	129
	Label	130
20	Capture	131
	Label	132
	Label	133
	Label	134
	Label	135
25	Capture	136
•	Label	137
	Label	138

Table 4 continued

	Probe Type	Sequence No.
5	Label	139
	Capture	140
	Label	141
	Label	142
	Label	143
10	Capture	144
	Label	145

The detection and capture probe HCV-specific segments, and their respective names as used in this assay were as follows.

Capture sequences are sequences numbered 119-122 and 141-144.

Detection sequences are sequences numbered 119-140.

the sequences substantially complementary to the HCV sequences, a 5' extension (B) which extension (B) is complementary to a segment of the second amplifier nucleic acid. The extension (B) sequence is identified in the Sequence Listing as Sequence No. 146, and is reproduced below.

AGGCATAGGACCCGTGTCTT

25

Each capture sequence contained, in addition to the sequences substantially complementary to HCV sequences, a sequence complementary to DNA bound to a solid phase. The sequence complementary to DNA bound to a solid support was carried downstream from the capture sequence. The sequence complementary to the DNA bound to the support is set forth as Sequence No. 147 and is reproduced below.

CTTCTTTGGAGAAAGTGGTG

Microtiter plates were prepared as follows. White Microlite 1 Removawell strips (polystyrene microtiter plates, 96 wells/plate) were purchased from Dynatech Inc.

Each well was filled with 200 µl 1 N HCl and incubated at room temperature for 15-20 min. The plates were then washed 4 times with 1X PBS and the wells aspirated to remove liquid. The wells were then filled with 200 µl 1 N NaOH and incubated at room temperature for 15-20 min. The plates were again washed 4 times with 1X PBS and the wells aspirated to remove liquid.

Poly(phe-lys) was purchased from Sigma Chemicals, Inc. This polypeptide has a 1:1 molar ratio of phe:lys and an average m.w. of 47,900 gm/mole. It has an average length of 309 amino acids and contains 155 amines/mole. A 1 mg/ml solution of the polypeptide was mixed with 2M NaCl/lx PBS to a final concentration of

10

15

20

0.1 mg/ml (pH 6.0). A volume of 200 μ l of this solution was added to each well. The plate was wrapped in plastic to prevent drying and incubated at 30°C overnight. The plate was then washed 4 times with 1X

PBS and the wells aspirated to remove liquid. The following procedure was used to couple the nucleic acid, a complementary sequence to Sequence No. 147, to the plates, hereinafter referred to as immobilized nucleic acid. Synthesis of immobilized nucleic acid having a sequence complementary to Sequence No. 133 was described in EPA 883096976. A quantity of 20 mg disuccinimidyl suberate was dissolved in 300 μ l dimethyl formamide (DMF). A quantity of 26 OD₂₆₀ units of immobilized nucleic acid was added to 100 µl coupling buffer (50 mM sodium phosphate, pH 7.8). The coupling mixture was then added to the DSS-DMF solution and stirred with a magnetic stirrer for 30 min. An NAP-25 column was equilibrated with 10 mM sodium phosphate, pH 6.5. The coupling mixture DSS-DMF solution was added to 2 ml 10 mM sodium phosphate, pH 6.5, at 4°C. The mixture was vortexed to mix and loaded onto the equilibrated NAP-25 column. DSS-activated immobilized nucleic acid DNA was eluted from the column with 3.5 ml 10 mM sodium phosphate, pH A quantity of 5.6 OD_{260} units of eluted DSS-activated immobilized nucleic acid DNA was added to

25 1500 ml 50 mM sodium phosphate, pH 7.8. A volume of 50

10

15

20

25

plates were incubated overnight. The plate was then washed 4 times with 1X PBS and the wells aspirated to remove liquid.

Final stripping of plates was accomplished as follows. A volume of 200 µl of 0.2N NaOH containing 0.5% (w/v) SDS was added to each well. The plate was wrapped in plastic and incubated at 65°C for 60 min. The plate was then washed 4 times with 1X PBS and the wells aspirated to remove liquid. The stripped plate was stored with desiccant beads at 2-8°C.

Serum samples to be assayed were analyzed using PCR followed by sequence analysis to determine the genotype.

Sample preparation consisted of delivering 50 µl of the serum sample and 150 µl P-K Buffer (2 mg/ml proteinase K in 53 mM Tris-HCl, pH 8.0/0.6 M NaCl/0.06 M sodium citrate/8 mM EDTA, pH 8.0/1.3%SDS/16µg/ml sonicated salmon sperm DNA/7% formamide/50 fmoles capture probes/160 fmoles detection probes) to each well. Plates were agitated to mix the contents in the well, covered and incubated for 16 hr at 62°C.

After a further 10 minute period at room temperature, the contents of each well were aspirated to remove all fluid, and the wells washed 2X with washing buffer (0.1% SDS/0.015 M NaCl/ 0.0015 M sodium citrate). The amplifier nucleic acid was then added to

each well (50 µl of 0.7 fmole/µl solution in 0..48 M NaCl/0.048 M sodium citrate/0.1% SDS/0.5% "blocking reagent" (Boehringer Mannheim, catalog No. 1096 176)). After covering the plates and agitating to mix the contents in the wells, the plates were incubated for 30 min. at 52°C.

5

10

25

After a further 10 min period at room temperature, the wells were washed as described above.

Alkaline phosphatase label nucleic acid, disclosed in EP 883096976, was then added to each well (50 µl/well of 2.66 fmoles/µl). After incubation at 52°C for 15 min., and 10 min. at room temperature, the wells were washed twice as above and then 3X with 0.015 M NaCl/0.0015 M sodium citrate.

15 An enzyme-triggered dioxetane (Schaap et al., Tet. Lett. (1987) 28:1159-1162 and EPA Pub. No. 0254051), obtained from Lumigen, Inc., was employed. A quantity of 50 μl Lumiphos 530 (Lumigen) was added to each well. The wells were tapped lightly so that the reagent would fall to the bottom and gently swirled to distribute the reagent evenly over the bottom. The wells were covered and incubated at 37°C for 20-40 min.

Plates were then read on a Dynatech ML 1000 luminometer. Output was given as the full integral of the light produced during the reaction.

The assay positively detected each of the serum samples, regardless of genotype.

IV. Expression of the Polypeptide Encoded in Sequences Defined by Differing Genotypes

HCV polypeptides encoded by a sequence within sequences 1-66 are expressed as a fusion polypeptide with superoxide dismutase (SOD). A cDNA carrying such sequences is subcloned into the expression vector psoDcfl (Steimer et al. 1986)).

First, DNA isolated from pSODcfl is treated with BamHI and EcoRI, and the following linker was ligated into the linear DNA created by the restriction enzymes:

GAT CCT GGA ATT CTG ATA AGA

CCT TAA GAC TAT TTT AA After cloning, the plasmid containing the insert is isolated.

Plasmid containing the insert is restricted with 15 EcoRI. The HCV cDNA is ligated into this EcoRI linearized plasmid DNA. The DNA mixture is used to transform E. coli strain D1210 (Sadler et al. (1980)). Polypeptides are isolated on gels.

20

25

5

10

Antigenicity of Polypeptides v.

The antigenicity of polypeptides formed in Section IV is evaluated in the following manner. Polyethylene pins arranged on a block in an 8 12 array (Coselco Mimetopes, Victoria, Australia) are prepared by placing the pins in a bath (20% v/v piperidine in dimethylformamide (DMF)) for 30 minutes at room

15

20

25

temperature. The pins are removed, washed in DMF for 5 minutes, then washed in methanol four times (2 min/wash). The pins are allowed to air dry for at least 10 minutes, then washed a final time in DMF (5Min). 1-Hydroxybenzotriazole (HOBt, 367 mg) is dissolved in DMF (80 μ L) for use in coupling Fmoc-protected polypeptides prepared in Section IV.

The protected amino acids are placed in micro-titer plate wells with HOBt, and the pin block placed over the plate, immersing the pins in the wells. The assembly is then sealed in a plastic bag and allowed to react at 25°C for 18 hours to couple the first amino acids to the pins. The block is then removed, and the pins washed with DMF (2 min.), MeOH (4 x, 2 min.), and again with DMF (2 min.) to clean and deprotect the bound amino acids. The procedure is repeated for each additional amino acid coupled, until all octamers are prepared.

The free N-termini are then acetylated to compensate for the free amide, as most of the epitopes are not found at the N-terminus and thus would not have the associated positive charge. Acetylation is accomplished by filling the wells of a microtiter plate with DMF/acetic anhydride/triethylamine (5:2:1 v/v/v) and allowing the pins to react in the wells for 90 minutes at 20°C. The pins are then washed with DMF (2

10

15

20

25

min.) and MeOH (4 \times , 2 min.), and air dried for at least 10 minutes.

The side chain protecting groups are removed by treating the pins with trifluoroacetic acid/phenol/dithioethane (95:2.5:1.5, v/v/v) in polypropylene bags for 4 hours at room temperature. The pins are then washed in dichloromethane (2 x, 2 min.), 5% di-isopropylethylamine/dichloromethane (2 x, 5 min.), dichloromethane (5 min.), and air-dried for at least 10 minutes. The pins are then washed in water (2 min.), MeOH (18 hours), dried in vacuo, and stored in sealed plastic bags over silica gel. IV.B.15.b Assay of Peptides.

Octamer-bearing pins are treated by sonicating for 30 minutes in a disruption buffer (1% sodium dodecylsulfate, 0.1% 2-mercaptoethanol, 0.1 M NaH2PO4) at 60°C. The pins are then immersed several times in water (60°C), followed by boiling MeOH (2 min.), and allowed to air dry.

The pins are then precoated for 1 hour at 25°C in microtiter wells containing 200 µL blocking buffer (1% ovalbumin, 1% BSA, 0.1% Tween, and 0.05% NaN3 in PBS), with agitation. The pins are then immersed in microtiter wells containing 175 µL antisera obtained from human patients diagnosed as having HCV and allowed to incubate at 4°C overnight. The formation of a complex between polyclonal antibodies of the serum and

the polypeptide initiates that the peptides give rise to an immune response in vivo. Such peptides are candidates for the development of vaccines.

Thus, this invention has been described and illustrated. It will be apparent to those skilled in the art that many variations and modifications can be made without departing from the purview of the appended claims and without departing from the teaching and scope of the present invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- 5 (i) APPLICANT: Tai-An Cha
 - (ii) TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS
- 10 (iii) NUMBER OF SEQUENCES: 147
 - (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
 - (B) STREET: 600 Atlantic Avenue
- 15 (C) CITY: Boston
 - (D) STATE: Massachusetts
 - (E) COUNTRY: USA
 - (F) ZIP: 02210
- 20 (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Diskette, 5.25 inch
 - (B) COMPUTER: IBM compatible
 - (C) OPERATING SYSTEM: MS-DOS Version 3.3
 - (D) SOFTWARE: WordPerfect 5.1

	(vi)	CURRENT APPLICATION DATA:
	•	(A) APPLICATION NUMBER: Not Available
		(B) FILING DATE: Not Available
		(C) CLASSIFICATION: Not Available
5		
	(vii)	PRIOR APPLICATION DATA:
	•	(A) APPLICATION NUMBER: 07/697,326
		(B) FILING DATE: 8 May 1991
10	(viii)	ATTORNEY/AGENT INFORMATION:
		(A) NAME: Janiuk, Anthony J.
	•	(B) REGISTRATION NUMBER: 29,809
		(C) REFERENCE/DOCKET NUMBER: C0772/7000
15	(ix)	TELECOMMUNICATION INFORMATION:
		(A) TELEPHONE: (617) 720-3500
		(B) TELEFAX: (617) 720-2441
		(C) TELEX: EZEKIEL
20	(2) INFORM	MATION FOR SEQ ID NO: 1:
	(i)	SEQUENCE CHARACTERISTICS:
		(A) LENGTH: 340 nucleotides
		(B) TYPE: nucleic acid
25		(C) STRANDEDNESS: single
		(D) TOPOLOGY: linear

- 54 -

	(ii) MOLECULE TYP	PE: DNA
	(vi) ORIGINAL SOU	JRCE: (ATCC # 40394) IDUAL ISOLATE: ns5hcv1
5	CTCCACAGTC ACTGAGA ATCTACCAAT GTTGTGA CCATCAAGTC CCTCACC TCTTACCAAT TCAAGGG TGCCGCGCGA GCGGCGT CCCTCACTTG CTACATC	SCRIPTION: SEQ ID NO: 1 GCG ACATCCGTAC GGAGGAGGCA CCT CGACCCCAA GCCCGCGTGG GAG AGGCTTTATG TTGGGGGCCC GGG AGAACTGCGG CTATCGCAGG ACT GACAACTAGC TGTGGTAACA CAAG GCCCGGGCAG CCTGTCGAGC CGCA CCATGCTCGT GTGTGGCGAC STGA AAGCGCGGGG GTCCAGGAGG 34
15	ACGCGGCGAG CCTGAGA	AGCC
-	(2) INFORMATION FOR SI	
20	(A) LENG (B) TYPE (C) STRA	HARACTERISTICS: TH: 340 nucleotides : nucleic acid NDEDNESS: single NLOGY: linear
25	(ii) MOLECULE I	TYPE: DNA

- 55 -

	•	(vi)	ORIG	INAL	SOURC	E:				
			(C)	INI	OIVIDU	AL ISC	LATE:	ns5i2	l	
		(xi)	SEQU	ENCE	DESCR	IPTION	I: SEQ	ID NO:	2	
5		CTCCA	CAGTC I	ACTGA	GAGCG	ACATO	CGTAC	GGAGGAG	GCA	4
		ATTTA	CCAAT (GTTGI	GACCT	GGACC	CCCAA	GCCCGCZ	ATGG	8
		CCATC	AAGTC (CTCA	CTGAG	AGGCI	TTATG	TCGGGG	SCCC.	12
		TCTTA	CCAAT :	CAAG	GGGGG	AGAAC	TGCGG	CTACCGC	CAGG	16
		TGCCG	CGCGA (3CGGC	GTACT	GACAA	CTAGC	TGTGGTA	AACA	20
10		CCCTC	ACTTG (TACA	TCAAG	GCCCG	IGGCAG	CCTGTCG	BAGC	24
		CGCAG	GCTC (AGGA	CTGCA	CCATG	CTTGT	GTGTGGC	GAC	280
		GACTT	AGTCG T	TATO	TGTGA	AAGTG	CGGGG	GTCCAGG	AGG	320
	-	ACGCG	GCGAG (CTGA	GAGCC	,				340
15	(2)	INFOR	MOITAN	FOR	SEQ II	O NO:	3:	·		
,		(i)	SEQUE	NCE	CHARAC	CTERIS	TICS:			
			(A)	LEN	GTH:	340 n	ucleot	tides		
			(B)	TYP	E: ni	cleic	acid			
20		•	(C)	STR	ANDEDI	VESS:	sing	le		
			(D)	TOP	OLOGY:	lin	ear			
		(ii)	MOLEC	ULE	TYPE:	DNA		·		
25		(vi)	ORIGI	NAL :	SOURCE	E:				
			(C)	ind	ividua	l iso	late:	ns5pt1		

		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3	
		CTCCACAGTC ACTGAGAGCG ACATCCGTAC GGAGGAGGCA	40
		ATCTACCAAT GTTGTGATCT GGACCCCCAA GCCCGCGTGG	80
		CCATCAAGTC CCTCACTGAG AGGCTTTACG TTGGGGGCCC	120
5		TCTTACCAAT TCAAGGGGGG AGAACTGCGG CTACCGCAGG	160
		TGCCGGGCGA GCGGCGTACT GACAACTAGC TGTGGTAATA	200
		CCCTCACTTG CTACATCAAG GCCCGGGCAG CCTGTCGAGC	240
		CGCAGGGCTC CGGGACTGCA CCATGCTCGT GTGTGGTGAC	280
		GACTTGGTCG TTATCTGTGA GAGTGCGGGG GTCCAGGAGG	320
10		ACGCGGCGAG CCTGAGAGCC	340
		•	
	(2)	INFORMATION FOR SEQ ID NO: 4	•
		(i) SEQUENCE CHARACTERISTICS:	
15		(A) LENGTH: 340 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
•		(D) TOPOLOGY: linear	
20		(ii) MOLECULE TYPE: DNA	
		A A A A A A A A A A A A A A A A A A A	
		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: ns5gm2	
05		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4	
25		CTCTACAGTC ACTGAGAACG ACATCCGTAC GGAGGAGGCA	40
		ATTTACCART GTTGTGACCT GGACCCCAA GCCCGCGTGG	80
		ATTITION AND THE STATE OF THE S	_

	_	CCATCAAGTC CCTCACTGAG AGGCTTTATG TTGGGGGCCC	120
	•	CCTTACCAAT TCAAGGGGGG AAAACTGCGG CTATCGCAGG	160
	•	TGCCGCGCGA GCGGCGTACT GACAACTAGC TGTGGTAACA	200
		CCCTCACTTG CTACATTAAG GCCCGGGCAG CCTGTCGAGC	240
_		CGCAGGGCTC CAGGACTGCA CCATGCTCGT GTGTGGCGAC	280
5		GACTTAGTCG TTATCTGTGA GAGTGCGGGA GTCCAGGAGG	320
•		ACGCGCGAA CTTGAGAGCC	340
		ACGCGGCGAA CIIGAGAGCC	
10	(2)	INFORMATION FOR SEQ ID NO: 5	
10		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 340 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
15	·	(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
		(vi) ORIGINAL SOURCE:	
20		(C) INDIVIDUAL ISOLATE: ns5us17	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5	4.5
		CTCCACAGTC ACTGAGAGCG ATATCCGTAC GGAGGAGGCA	40
		ATCTACCAGT GTTGTGACCT GGACCCCCAA GCCCGCGTGG	80
25	٠	CCATCAAGTC CCTCACCGAG AGGCTTTATG TCGGGGGCCC	120
		TCTTACCAAT TCAAGGGGGG AAAACTGCGG CTATCGCAGG	160
		TGCCGCGCAA GCGGCGTACT GACAACTAGC TGTGGTAACA	200

		CCCTCACTTG TTACATCAAG GCCCAAGCAG CCTGTCGAGC	240
		CGCAGGGCTC CGGGACTGCA CCATGCTCGT GTGTGGCGAC	280
		GACTTAGTCG TTATCTGTGA AAGTCAGGGA GTCCAGGAGG	320
		ATGCAGCGAA CCTGAGAGCC	340
5			
	(2)	INFORMATION FOR SEQ ID NO: 6	
•	-	(i) SEQUENCE CHARACTERISTICS:	
: ·		(A) LENGTH: 340 nucleotides	
10		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
15		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: ns5sp2	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6	
20		CTCTACAGTC ACTGAGAGCG ATATCCGTAC GGAGGAGGCA	40
		ATCTACCAAT GTTGTGACCT GGACCCCGAA GCCCGTGTGG	80
		CCATCAAGTC CCTCACTGAG AGGCTTTATG TTGGGGGCCC	120
		TCTTACCAAT TCAAGGGGG AGAACTGCGG CTACCGCAGG	160
		TGCCGCGCAA GCGGCGTACT GACGACTAGC TGTGGTAATA	200
25		CCCTCACTTG TTACATCAAG GCCCGGGCAG CCTGTCGAGC	240
4 3		CGCAGGGCTC CAGGACTGCA CCATGCTCGT GTGTGGCGAC	280
		Correcto rue cue cue cue cue cue cue cue cue cue c	

- 59 -

		GACCTAGTCG TTATCTGCGA AAGTGCGGGG GTCCAGGA	.GG 320
		ACGCGGCGAG CCTGAGAGCC	. 340
5	(2)	INFORMATION FOR SEQ ID NO: 7	
•		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 340 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
10		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
		(vi) ORIGINAL SOURCE:	
15		(C) INDIVIDUAL ISOLATE: ns5j1	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7	
		CTCCACAGTC ACTGAGAATG ACACCCGTGT TGAGGAGT	CA 40
		ATTTACCAAT GTTGTGACTT GGCCCCCGAA GCCAGACA	GG 80
20		CCATAAGGTC GCTCACAGAG CGGCTCTATG TCGGGGGT	CC 120
		TATGACTAAC TCCAAAGGGC AGAACTGCGG CTATCGCC	GG 160
		TGCCGCGCGA GCGGCGTGCT GACGACTAGC TGCGGTAA	TA 200
		CCCTCACATG CTACCTGAAG GCCACAGCGG CCTGTCGA	GC 240
		TGCCAAGCTC CAGGACTGCA CGATGCTCGT GAACGGAG	AC 280
25		GACCTTGTCG TTATCTGTGA AAGCGCGGGG AACCAAGA	GG 320
		ACGCGGCAAG CCTACGAGCC	340

	(2)	INFORMATION FOR SEQ ID NO: 8	
5		(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 340 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
10		<pre>(vi) ORIGINAL SOURCE: (C) INDIVIDUAL ISOLATE: ns5kl</pre>	
15		(XI) SEQUENCE DESCRIPTION: SEQ ID NO: 8 CTCAACGGTC ACTGAGAATG ACATCCGTGT TGAGGAGTCA ATTTACCAAA GTTGTGACTT GGCCCCCGAG GCCAGACAAG CCATAAGGTC GCTCACAGAG CGGCTTTACA TCGGGGGCCC CCTGACTAAT TCAAAAGGGC AGAACTGCGG CTATCGCCGA TGCCGCGCCA GCGGTGTGCT GACGACTAGC TGCGGTAATA CCCTCACATG TTACTTGAAG GCCACTGCGG CCTGTAGAGC TGCGAAGCTC CAGGACTGCA CGATGCTCGT GTGCGGAGAC GACCTTGTCG TTATCTGTGA AAGCGCGGGA ACCCAGGAGG ATGCGGCGAG CCTACGAGTC	40 80 120 160 240 240 320 34
25	(2) INFORMATION FOR SEQ ID NO: 9	

		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 340 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
5		(D) TOPOLOGY: linear	
	,	(ii) MOLECULE TYPE: DNA	
		(vi) ORIGINAL SOURCE:	
10		(C) INDIVIDUAL ISOLATE: ns5kl.1	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9	
		CTCAACGGTC ACCGAGAATG ACATCCGTCT TOTAL	40
		ATTATCAAT GIIGIGCCII GGCCCCCCIIC COLLECTION	В0
15		CCATAAGGTC GCTCACAGAG CGGCTTAIN 2000000000000000000000000000000000000	20
		CCTGACCAAT TCAAAGGGGC AGAACTGCGG TTATCGCCGG 1	60
		TGCCGCGCCA GCGGCGTACT GACGACGTOG TGGGCGCCA	00
		CCCTTACATG TTACTTGAAG GCCTCTGCAG CCTGTCGAGC 2-	40
		CGCGAAGCTC CAGGACTGCA CGATGCTCGT GTGTGGGGAC 2	80
20		GACCTTGTCG TTATCTGTGA AAGCGCGGGA ACCCAGGAGG 3:	20
		ACGCGGCGAA CCTACGAGTC 3	40
	(2)	INFORMATION FOR SEQ ID NO: 10	
25		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 340 nucleotides	
		(R) TYPE: nucleic acid	

		(C) STRANDEDNESS: Single	
		(D) TOPOLOGY: linear	
_		(ii) MOLECULE TYPE: DNA	
5		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: ns5gh6	
	•	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10	
10		CTCAACGGTC ACTGAGAGTG ACATCCGTGT CGAGGAGTCG	40
		ATTTACCAAT GTTGTGACTT GGCCCCCGAA GCCAGGCAGG	80
		CCATAAGGTC GCTCACCGAG CGACTTTATA TCGGGGGCCC	
		CCTGACTAAT TCAAAAGGGC AGAACTGCGG TTATCGCCGG	160
		TGCCGCGCGA GCGGCGTGCT GACGACTAGC TGCGGTAATA	200
15		CCCTCACATG TTACTTGAAG GCCTCTGCAG CCTGTCGAGC	240
		TGCAAAGCTC CAGGACTGCA CGATGCTCGT GAACGGGGAC	280
		GACCTTGTCG TTATCTGCGA GAGCGCGGGA ACCCAAGAGG	320
		ACGCGCGAG CCTACGAGTC	340
		ACGCGGCGAG COLINOLIULO	
20	(2)	INFORMATION FOR SEQ ID NO: 11	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 340 nucleotides	
		(B) TYPE: nucleic acid	
25		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	

		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
			(C) INDIVIDUAL ISOLATE: ns5spl	
5				
			SEQUENCE DESCRIPTION: SEQ ID NO: 11	
			AGTC ACTGAGAGTG ACATCCGTGT TGAGGAGTCA	40
			CAAT GTTGTGACTT GGCCCCCGAA GCCAGACAGG	80
	:	CTATAA	AGGTC GCTCACAGAG CGGCTGTACA TCGGGGGTCC	
10		CCTGAC	TAAT TCAAAAGGGC AGAACTGCGG CTATCGCCGG	160
		TGCCGC	GCAA GCGGCGTGCT GACGACTAGC TGCGGTAACA	200
		CCCTCA	ACATG TTACTTGAAG GCCTCTGCGG CCTGTCGAGC	240
		TGCGAA	AGCTC CAGGACTGCA CGATGCTCGT GTGCGGTGAC	280
		GACCTI	GTCG TTATCTGTGA GAGCGCGGGA ACCCAAGAGG	320
15		ACGCGG	GCGAG CCTACGAGTC	340
	(2)	INFORM	MATION FOR SEQ ID NO: 12	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 340 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	

		(C) individual isolate: ns5sp3	
5		(XI) SEQUENCE DESCRIPTION: SEQ ID NO: 12 CTCAACAGTC ACTGAGAGTG ACATCCGTGT TGAGGAGTCA ATCTACCAAT GTTGTGACTT GGCCCCCGAA GCCAGACAGG CTATAAGGTC GCTCACAGAG CGGCTTTACA TCGGGGGTCC CCTGACTAAT TCAAAAGGGC AGAACTGCGG CTATCGCCGG TGCCGCGCAA GCGGCGTGCT GACGACTAGC TGCGGTAATA CCCTCACATG TTACCTGAAG GCCAGTGCGG CCTGTCGAGC	160 200
LO		TGCGAAGCTC CAGGACTGCA CAATGCTCGT GTGCGGTGAC	280 320 340
15	(2)	<pre>INFORMATION FOR SEQ ID NO: 13 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 340 nucleotides (B) TYPE: nucleic acid</pre>	
20		(C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA	
25		<pre>(vi) ORIGINAL SOURCE:</pre>	

		CTCAACCGTC ACTGAGAGAG ACATCAGAAC TGAGGAGTCC	40
		ATATACCGAG CCTGCTCCCT GCCTGAGGAG GCTCACATTG	80
		CCATACACTC GCTGACTGAG AGGCTCTACG TGGGAGGGCC	120
		CATGTTCAAC AGCAAGGGCC AGACCTGCGG GTACAGGCGT	160
5		TGCCGCGCCA GCGGGGTGCT CACCACTAGC ATGGGGAACA	200
		CCATCACATG CTATGTAAAA GCCCTAGCGG CTTGCAAGGC	240
		TGCAGGGATA GTTGCACCCT CAATGCTGGT ATGCGGCGAC	280
		GACTTAGTTG TCATCTCAGA AAGCCAGGGG ACTGAGGAGG	320
		ACGAGCGGAA CCTGAGAGCT	340
10			
	(2)	INFORMATION FOR SEQ ID NO: 14	
		(i) SEQUENCE CHARACTERISTICS:	-
		(A) LENGTH: 340 nucleotides	
15		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
20			
		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: ns5arg8	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14	
25		CTCTACAGTC ACGTAAAAGG ACATCACATC CTAGGAGTCC	40
		ATCTACCAGT CCTGTTCACT GCCCGAGGAG GCTCGAACTG	80
		CTATACACTC ACTGACTGAG AGACTATACG TAGGGGGGCC	120

		CATGACAAAC AGCAAGGGCC AATCCTGCGG GTACAGGCGT TGCCGCGCGA GCGCAGTGCT CACCACCAGC ATGGGCAACA	200
		TGCCGCGCGA GCGCAGTOOT GTGCCGCGACCCCCGACCCCGCGCGCGCGCGGCGGCGGCGGC	240
		CACTCACGTG CTACGTASTS CONTROL GTGCGGTGAC CGCGGGGATT GTTGCTCCCA CCATGCTGGT GTGCGGTGAC	280
		CGCGGGGATT GTTGCICCCA GGILLOGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	320
5		GACCTGGTCG TCATCICAGA GAGTC	340
		ACGAGCAGAA CCTGAGAGTC	
	(2)	INFORMATION FOR SEQ ID NO: 15	
10	·	(i) SEQUENCE CHARACTERISTICS:	
TO		(A) LENGTH: 340 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
15		(ii) MOLECULE TYPE: DNA	
		<pre>(vi) ORIGINAL SOURCE: (C) INDIVIDUAL ISOLATE: ns5il0</pre>	٠
20		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15	
		(xi) SEQUENCE DESCRIPTION. 529 25 CTCTACAGTC ACAGAGGGGGGGGGGGGGGGGGGGGGGG	40
		ATCTATCTGT CCTGCTCACT GCCTGAGGAG GCCCGAACTG	80
		ATCTATCTGT CCTGCTGAG AGACTGTACG TAGGGGGGCC	120
		CTATACACTC ACTGACTOR CATGACAGCGT CATGACAAAC AGCAAGGGGC AATCCTGCGG GTACAGGCGT	160
25		TGCCGCGCGA GCGGAGTGCT CACCACCAGC ATGGGCAACA	200
		TGCCGCGCGA GCGGAGTGTAACGC CGCTCACGTG CTACGTGAAA GCCAGAGCGG CGTGTAACGC	240
		CGCTUACGIG CIACCITITE	

		CGCGGGCATT GTTGCTCCCA CCATGTTGGT GTGCGGCGAC	280
		GACCTGGTTG TCATCTCAGA GAGTCAGGGG GTCGAGGAAG	320
		ATGAGCGGAA CCTGAGAGTC	340
5	(2)	INFORMATION FOR SEQ ID NO: 16	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 340 nucleotides	
		(B) TYPE: nucleic acid	
10		(C) STRANDEDNESS: single	
	-	(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
15		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: ns5arg6	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16	
		CTCTACAGTC ACGGAGAGGG ACATCAGAAC CGAGGAGTCC	40
20		ATCTATCTGT CCTGTTCACT GCCTGAGGAG GCTCGAACTG	80
		CCATACACTC ACTGACTGAG AGGCTGTACG TAGGGGGGCC	120
		CATGACAAAC AGCAAAGGGC AATCCTGCGG GTACAGGCGT	160
		TGCCGCGCGA GCGGAGTGCT CACCACCAGC ATGGGTAACA	200
		CACTCACGTG CTACGTGAAA GCTAAAGCGG CATGTAACGC	240
25		CGCGGCCATT GTTGCCCCCA CCATGTTGGT GTGCGGCGAC	280
		GACCTAGTCG TCATCTCAGA GAGTCAAGGG GTCGAGGAGG	320
		ATGAGCGAAA CCTGAGAGCT	340

	(2)	INFORMATION FOR SEQ 12 No. 1	
5		(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 340 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
10		(ii) MOLECULE TYPE: DNA(vi) ORIGINAL SOURCE:(C) INDIVIDUAL ISOLATE: ns5k2b	
15		CTATCCACTC GCTCACTGAG AGACTCTACG TAGGACCCTC	_
20		TGCCGCGCCA GCGGGGTCTT CACCACCAGC ATGGGGAAAGC 2 CCATGACATG CTACATCAAA GCCCTTGCAG CGTGCAAAGC 2 TGCAGGGATC GTGGACCCTA TCATGCTGGT GTGTGGAGAC 2 CACCTGGTCG TCATCTCGGA GAGCGAAGGT AACGAGGAGG 3	40 80 20
25	(2)	INFORMATION FOR SEQ ID NO: 18 (i) SEQUENCE CHARACTERISTICS:	

		(A) LENGTH: 340 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
5			
		(ii) MOLECULE TYPE: DNA	•
		(!) OPTOTAL COMPORT	
		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: ns5sa283	
10			
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18	
			40
		ATTTACCAAT CATTGTACTT GCAGCCTGAG GCGCGTGTGG	80
		CAATACGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCCC 1	20
15		CATGTATAAC AGCAAGGGGC AACAATGTGG TTATCGTAGA 1	60
		TGCCGCGCCA GCGGCGTCTT CACCACTAGT ATGGGCAACA 2	00
		CCATGACGTG CTACATTAAG GCTTTAGCCT CCTGTAGAGC 2	40
		CGCAAAGCTC CAGGACTGCA CGCTCCTGGT GTGTGGTGAT 3	20
		GATAAAGCGA CCTGAGAGCC 3	40
20			
	(2)	INFORMATION FOR SEQ ID NO: 19	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 340 nucleotides	
25		(B) TYPE: nucleic acid	
23		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(n) intomosi. Timesi	

		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
			(C) INDIVIDUAL ISOLATE: ns5sa156	
5				
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 19	
			CCGTT ACCGAACATG ACATAATGAC TGAAGAGTCC	40
		ATTTAC	CCAAT CATTGTACTT GCAGCCTGAG GCACGCGCGG	80
			CGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCCC	
10		CATGTA	ATAAC AGCAAGGGGC AACAATGTGG TTACCGTAGA	160
		TGCCGC	CGCCA GCGGCGTCTT CACCACCAGT ATGGGCAACA	200
		CCATGA	ACGTG CTACATCAAG GCTTCAGCCG CCTGTAGAGC	240
•		TGCAAA	AGCTC CAGGACTGCA CGCTCCTGGT GTGTGGTGTG	280
		ACCTTG	GTGG CCATTTGCGA GAGCCAAGGG ACGCACGAGG	320
15		ATGAAG	SCGTG CCTGAGAGTC	340
	(2)	INFORM	MATION FOR SEQ ID NO: 20	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 340 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	`
25		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	

(C) INDIVIDUAL ISOLATE: ns5il1

		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20	
		CTCTACTGTC ACTGAACAGG ACATCAGGGT GGAAGAGGAG	40
5		ATATACCAGT GCTGTAACCT TGAACCGGAG GCCAGGAAAG	80
.		TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC	120
		TATGTTCAAC AGCAAGGGGG CCCAGTGTGG TTATCGCCGT	160
		TGCCGTGCTA GTGGAGTCCT GCCTACCAGC TTCGGCAACA	200
			240
		CAATCACTTG TTACATCAAG GCTAGAGCGG CTTCGAAGGC	000
10		CGCAGGCCTC CGGAACCCGG ACTTTCTTGT CTGCGGAGAT	280
•		GATCTGGTCG TGGTGGCTGA GAGTGATGGC GTCGACGAGG	320
		ATAGAGCAGC CCTGAGAGCC	340
		ALNO.	
	(2)	INFORMATION FOR SEQ ID NO: 21	
15	\		
77		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 340 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
20		(D) TOPOLOGY: linear	

(ii) MOLECULE TYPE: DNA

(C)

(vi)

(xi)

25

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: ns5i4

SEQUENCE DESCRIPTION: SEQ ID NO: 21

		CTCGACTGTC ACTGAACAGG ACATCAGGGT GGAAGAGGGG	40
		ATATACCAAT GCTGTAACCT TGAACCGGAG GCCAGGAAAG	80
		TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC	120
		TATGTTCAAT AGCAAGGGGG CCCAGTGTGG TTATCGCCGT	160
5		TGCCGTGCTA GTGGAGTTCT GCCTACCAGC TTCGGCAACA	200
		CAATCACTTG TTACATCAAG GCTAGAGCGG CTGCGAAGGC	240
		CGCAGGGCTC CGGACCCCGG ACTTTCTCGT CTGCGGAGAT	280
		GATCTGGTTG TGGTGGCTGA GAGTGATGGC GTCGACGAGG	320
		ATAGAACAGC CCTGCGAGCC	340
10		•	
	(2)	INFORMATION FOR SEQ ID NO: 22	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 340 nucleotides	
15		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	•
••		(ii) MOLECULE TYPE: DNA	
20		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: ns5gh8	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22	•
25		CTCAACTGTC ACTGAACAGG ACATCAGGGT GGAAGAGGAG	40
		ATATACCAAT GCTGTAACCT TGAACCGGAG GCCAGGAAAG	80
		MANUSCOTT COTCACGGAA CGGCTTTACT GCGGGGGCCCC	120

5		CAATCACTTG TTACATCAAA GCTAGAGCGG CTGCCGAAGAT CGCAGGCCTC CGGAACCCGG ACTTTCTTGT CTGCGGAGAT CGCAGGCCTC TGGTGGCTGA GAGTGATGGC GTCAATGAGG 3:	-
	(2)	INFORMATION FOR SEQ ID NO: 23	
10		 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 100 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
15		(ii) MOLECULE TYPE: DNA	
		(vi) ORIGINAL SOURCE: (ATCC # 40394) (C) INDIVIDUAL ISOLATE: hcvl	
20		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23 GACGGCGTTG GTAATGGCTC AGCTGCTCCG GATCCCACAA GCCATCTTGG ACATGATCGC TGGTGCTCAC TGGGGAGTCC TGGCGGGCAT AGCGTATTTC	40 80 100
25	(2) INFORMATION FOR SEQ ID NO: 24	

		(1)	SEQUENCE CHARACTERISTICS:	
		÷	(A) LENGTH: 100 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
5			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
10			(C) INDIVIDUAL ISOLATE: US5	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 24	
		GACGG	CGTTG GTGGTAGCTC AGGTACTCCG GATCCCACAA	40
		GCCAT	CATGG ACATGATCGC TGGAGCCCAC TGGGGAGTCC	80
15		TGGCG	GCAT AGCGTATTTC	100
	(2)	INFOR	MATION FOR SEQ ID NO: 25	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 100 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	

- 75 -

		(C) INDIVIDUAL IS	OLATE: AUSS
5		(xi) SEQUENCE DESCRIPTION AACGGCGCTG GTAGTAGCTC AGCTGCTGG ACATGATCGC TGGTTAGCGGGCAT AGCGTATTTT	GCTCAG GGTCCCGCAA 40
	(2)	INFORMATION FOR SEQ ID NO:	26
10		(i) SEQUENCE CHARACTERI (A) LENGTH: 100 r (B) TYPE: nucleic (C) STRANDEDNESS: (D) TOPOLOGY: lin	ucleotides : acid : single
15		(ii) MOLECULE TYPE: DNA	4
•		(vi) ORIGINAL SOURCE: (C) INDIVIDUAL IS	SOLATE: US4
20		(xi) SEQUENCE DESCRIPTION GACAGCCCTA GTGGTATCGC AGT GCCGTCATGG ATATGGTGGC GGG TGGCGGGCCT TGCCTACTAT	TACTCCG GATCCCACAA 40
25	(2)	INFORMATION FOR SEQ ID NO	: 27

		(i)	SEQUENCE CHARACTERISTICS:	
		*	(A) LENGTH: 100 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
5			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	-
		(vi)	ORIGINAL SOURCE:	
10			(C) INDIVIDUAL ISOLATE: ARG2	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 27	
			CCTA GTGGTGTCGC AGTTACTCCG GATCCCACAA	40
			CGTGG ACATGGTGGC GGGGGCCCAC TGGGGAGTCC	80
15			GGCCT TGCTTACTAT	100
	(2)	INFOR	ATION FOR SEQ ID NO: 28	٠
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 100 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	

- 77 -

			(C) INDIVIDUAL ISOLATE: 115	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 28	
		GGCAG	CCCTA GTGGTGTCGC AGTTACTCCG GATCCCGCA	A 40
5		GCTGT	CGTGG ACATGGTGGC GGGGGCCCAC TGGGGAATC	C 80
		TAGCG	GGTCT TGCCTACTAT	100
	(2)	INFOR	MATION FOR SEQ ID NO: 29	
L O	·	(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 100 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
15		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
20			(C) INDIVIDUAL ISOLATE: GH8	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 29	
		TGTGG	GTATG GTGGTGGCGC ACGTCCTGCG TTTGCCCCA(3 40
		ACCTT	GTTCG ACATAATAGC CGGGGCCCAT TGGGGCATC	08 T
		TGGCG	GGCTT GGCCTATTAC	100
25				
	121	THEODI	MATTON FOR SEC ID NO. 30	

		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 100 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
5			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
10			(C) INDIVIDUAL ISOLATE: 14	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 30	
		TGTGG	STATE GTGGTAGCAC ACGTCCTGCG TCTGCCCCAG	40
		ACCTT	STTCG ACATAATAGC CGGGGCCCAT TGGGGCATCT	80
15		TGGCA	GCCT AGCCTATTAC	100
	(2)	INFORM	MATION FOR SEQ ID NO: 31	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 100 nucleotides	•
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	

	`.		(C) INDIVIDUAL ISOLATE: 111	
5		TGTGGG ACCTTG	SEQUENCE DESCRIPTION: SEQ ID NO: 31 TATE GTEGTEGCEC AAGTCCTECE TTTECCCCAE TTTCE ACGTECTACC CEGEGCCCAT TEGEGCATCT EGCCT EGCCTATTAC	40 80 100
	(2)	INFORM	NATION FOR SEQ ID NO: 32	
10		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 100 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
15		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE: (C) INDIVIDUAL ISOLATE: I10	
20		TACCA:	SEQUENCE DESCRIPTION: SEQ ID NO: 32 CTATG CTCCTGGCAT ACTTGGTGCG CATCCCGGAG CCTGG ACATTATCAC GGGAGGACAC TGGGGCGTGA GGCCT GGCTTATTTC	40 80 100
25	(2)	INFOR	MATION FOR SEQ ID NO: 33	

		(i)·	SEQUENCE CHARACTERISTICS:	
		•	(A) LENGTH: 252 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
5			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE: (ATCC # 40394)	
10			(C) INDIVIDUAL ISOLATE: hcvl	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 33	
			PATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC	40
		CGGGA	FAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
15		GGAAT'	PGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	120
			ATGCC TGGAGATTTG GGCGTGCCCC CGCAAGACTG	
			CGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCT	SATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	
		AGACC	STGCA CC	252
20				
	(2)	INFOR	MATION FOR SEQ ID NO: 34	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 252 nucleotides	
25			(B) TYPE: nucleic acid	
	•		(C) STRANDEDNESS: single	
			(n) TOPOLOGY: linear	

		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
		•	(C) INDIVIDUAL ISOLATE: us5	
5				
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 34	
		GTTAGI	TATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC	40
		CGGGAG	SAGCC ATAGTEGTUT GUGAACCOG TOAGTHAID	80
		GGAATI	IGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	120
10		GCTCA	ATGCC TGGAGATTTG GGCGTGCCCC CGCAAGACTG	160
		CTAGCO	CGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		ጥርርርጥር	GATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
			GTGCA CC	252
15	(2)	INFOR	MATION FOR SEQ ID NO: 35	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 252 nucleotides	
			(B) TYPE: nucleic acid	
20			(C) STRANDEDNESS: single	•
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
25		(vi)	ORIGINAL SOURCE:	
		-	(C) INDIVIDUAL ISOLATE: ausl	

		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35	
		GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC	40
		CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
		GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	120
5		GCTCAATGCC TGGAGATTTG GGCACGCCCC CGCAAGATCA	160
		CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
		AGACCGTGCA CC	252
10	(2)	INFORMATION FOR SEQ ID NO: 36	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 252 nucleotides	
		(B) TYPE: nucleic acid	
15		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
20		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: sp2	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36	
		GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC	40
25		CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
		GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATAAACCC	120
		GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG	160

	٠	CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
		AGACCGTGCA CC	252
5	(2)	INFORMATION FOR SEQ ID NO: 37	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 252 nucleotides	
		(B) TYPE: nucleic acid	
10		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
	••,	(ii) MOLECULE TYPE: DNA	
15		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: gm2	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37	
		GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC	40
20		CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
	•	GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	120
		GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCAAGACTG	160
		CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
25	** *	AGACCGTGCA CC	252
	(2)	INFORMATION FOR SEQ ID NO: 38	

		(1)	SEQUENCE CHARACTERISTICS.	
		•	(A) LENGTH: 252 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
5			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
10	•	-	(C) INDIVIDUAL ISOLATE: i21	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 38	
			TGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC	40
		CGGGAGA	GCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
15		GGAATTG	CCA GGACGACCGG GTCCTTTCTT GGATAAACCC	120
		GCTCAAT	GCC TGGAGATTTG GGCGTGCCCC CGCAAGACTG	160
		CTAGCCG	AGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTGA	TAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
		AGACCGT	GCA CC	252
20				
	(2)	INFORMA	TION FOR SEQ ID NO: 39	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 252 nucleotides	
25			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	

		(ii)	MOLECU	LE TYPE:	DNA			
		(vi)	ORIGIN	AL SOURC	Œ:			
		• • • • •		INDIVIDU		LATE:	us4	
5								
		(xi)	SEQUEN	CE DESCR	RIPTION	: SEQ	ID NO: 39	
•		GTTAG:	TATGA GT	STCGTGCA	GCCTC	CAGGA	CCCCCCTCC	40
		CGGGA	SAGCC ATA	AGTGGTCI	GCGGA	ACCGG	TGAGTACACC	80
		GGAAT'	rgcca ggi	ACGACCGG	GTCCT	TTCTT	GGATCAACCC	120
10		GCTCA	ATGCC TG	SAGATTTO	GGCGT	GCCCC	CGCGAGACTG	160
		CTAGC	GAGT AG	rgttgggi	CGCGA	AAGGC	CTTGTGGTAC	200
		TGCCT	SATAG GG	CCTTGC	AGTGC	CCCGG	GAGGTCTCGT	240
		AGACC	STGCA CC					252
15	(2)	INFOR	IATION FO	OR SEQ I	D NO:	40		
	-	(i)	SEQUEN	CE CHARA	CTERIS	TICS:		
			(A) I	LENGTH:	252 nu	cleot	ides	
			(B) 3	TYPE: nu	cleic	acid		
20			(C) S	TRANDEL	NESS:	sing:	le	
			(D) ?	COPOLOGY	: line	ar		
		(ii)	MOLECUI	LE TYPE:	DNA			
25		(vi)	ORIGINA	AL SOURC	Œ:			
			(c) 1	NDIVIDU	AL ISO	LATE:	jhl	

	•	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40	
	•	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC	40
	i	CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
		GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	120
5		GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG	160
		CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
		AGACCGTGCA TC	252
10	(2)	INFORMATION FOR SEQ ID NO: 41	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 252 nucleotides	
		(B) TYPE: nucleic acid	
15		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
20		(vi) ORIGINAL SOURCE:	
20		(C) INDIVIDUAL ISOLATE: nac5	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41	
		GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC	40
25		CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
		GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	120
		GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG	160

- 87 -

		CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
		AGACCGTGCA CC	252
5	(2)	INFORMATION FOR SEQ ID NO: 42	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 252 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
10		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	,
		(vi) ORIGINAL SOURCE:	
15		(C) INDIVIDUAL ISOLATE: arg2	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42	•
		GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC	40
		CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
20		GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	120
		GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG	160
		CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
		AGACCGTGCA CC	252
25			
	121	INFORMATION FOR CEO ID NO. 42	

		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 252 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
5			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
10			(C) INDIVIDUAL ISOLATE: spl	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 43	
			TATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC	40
			GAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
15			TGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	120
			ATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG	
			CGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	
		TGCCT	GATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
		AGACC	GTGCA CC	252
20			•	
	(2)	INFOR	MATION FOR SEQ ID NO: 44	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 252 nucleotides	
25			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(n) TOPOLOGY: linear	

	(ii)	MOLEC	ULE TYPE: DNA	
	(vi)	ORIGI	NAL SOURCE:	
		(C)	INDIVIDUAL ISOLATE:	ahl
5			•	•
	(xi)	SEQUE	NCE DESCRIPTION: SEQ II) NO: 44
	GTTAG	TATGA G	IGTCGTGCA GCCTCCAGGA CO	CCCCCTCC 40
	CGGGA	GAGCC A	PAGTGGTCT GCGGAACCGG TO	GAGTACACC 80
	GGAAT	TGCCA G	SACGACCGG GTCCTTTCTT GG	SATCAACCC 120
10	GCTCA	ATGCC T	GGAGATTTG GGCGTGCCCC CG	CGAGACTG 160
	CTAGC	CGAGT A	STGTTGGGT CGCGAAAGGC C1	TTGTGGTAC 200
	TGCCT	GATAG G	STGCTTGCG AGTGCCCCGG GA	AGGTCTCGT 240
	AGACC	GTGCA C	3	252
15 (2) INFOR	MATION	FOR SEQ ID NO: 45	
٠.				
	(i)		NCE CHARACTERISTICS:	
			LENGTH: 252 nucleotide	!S
		(B)	TYPE: nucleic acid	v
20		(C)	STRANDEDNESS: single	
		(D)	TOPOLOGY: linear	
				·
	(11)	MOLEC	LE TYPE: DNA	,
25	(vi)	ORIGIN	IAL SOURCE:	
	\ = /		00011001	

	(xī)	SEQUENCE DESCRIPTION: SEQ ID NO. 43	
	GTTAG!	TATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC	40
	CGGGA	GAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
	GGAAT'	TGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	120
	GCTCA	ATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG	160
	CTAGC	CGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
	TGCCT	GATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
	AGACC	GTGCA CC	252
(2)	INFOR	MATION FOR SEQ ID NO: 46	
	(i)		
	-	(A) LENGTH: 252 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(D) 10101001. 111011	•
	(44)		·
	(ii)	MOLECULE TYPE: DNA	·
	(ii) (vi)	MOLECULE TYPE: DNA	·
	(2)	CGGGA GGAAT GCTCA CTAGC TGCCT AGACC	CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACCC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT AGACCGTGCA CC (2) INFORMATION FOR SEQ ID NO: 46 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 252 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single

		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46	
		GCTAGTATCA GTGTCGTACA GCCTCCAGGC CCCCCCTCC	40
		CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
		GGAATTGCCG GGAAGACTGG GTCCTTTCTT GGATAAACCC	120
5		ACTCTATGCC CGGCCATTTG GGCGTGCCCC CGCAAGACTG	160
		CTAGCCGAGT AGCGTTGGGT TGCGAAAGGC CTTGTGGTAC	200
		TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
		AGACCGTGCA TC	252
10	(2)	INFORMATION FOR SEQ ID NO: 47	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 252 nucleotides	
		(B) TYPE: nucleic acid	
15		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
20		(vi) ORIGINAL SOURCE:	
20		(C) INDIVIDUAL ISOLATE: arg6	
		(C) INDIVIDUAL ISOLATE. algo	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47	
		GTTAGTATGA GTCTCGTACA GCCTCCAGGC CCCCCCTCC	40
25		CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
		GGAATTGCTG GGAAGACTGG GTCCTTTCTT GGATAAACCC	120
		ACTCTATGCC CAGCCATTTG GGCGTGCCCC CGCAAGACTG	160

		CTAGCCGAGT AGCGTTGGGT TGCGAAAGGC CTTGTGGTAC	200
			240
			252
_	(0)	INFORMATION FOR SEQ ID NO: 48	
5	(2)		
		(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 252 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
10		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
	•	(vi) ORIGINAL SOURCE:	
15		(C) INDIVIDUAL ISOLATE: s21	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48	
		GTTAGTACGA GTGTCGTGCA GCCTCCAGGA CTCCCCCTCC	40
20		CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
20			120
			160
			200
,		-	240
		\cdot	252
25		AGACCGTGCA AC	
	(2)	INFORMATION FOR SEQ ID NO: 49	

		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 252 nucleotides	
			(B) TYPE: nucleic acid	
5			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
10		(vi)	ORIGINAL SOURCE:	
			(C) INDIVIDUAL ISOLATE: gj61329	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 49	
15		GTTAGT	ACGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC	40
		CGGGAG	SAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
		GGAATC	GCTG GGGTGACCGG GTCCTTTCTT GGAGTAACCC	120
		GCTCAA	ATACC CAGAAATTTG GGCGTGCCCC CGCGAGATCA	160
		CTAGCC	CGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
20		TGCCTG	ATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
		AGACCG'	STGCA AC	252
	(2)	INFORM	ATION FOR SEQ ID NO: 50	
25		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 180 nucleotides	

			(B) TYPE: nucleic acid	
	:	•	(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
5		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
			(C) INDIVIDUAL ISOLATE: sa3	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 50	
10				4.0
			TATGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC	40
			GAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	
			TGCCG GGATGACCGG GTCCTTTCTT GGATAAACCC	
		GCTCA	ATGCC CGGAGATTTG GGCGTGCCCC CGCGAGACTG	160
15		CTAGC	CGAGT AGTGTTGGGT	180
	(2)	INFOR	MATION FOR SEQ ID NO: 51	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 180 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(vri)	ORIGINAL SOURCE:	

- 95 -

			(C)	IN	DIVIDU	AL ISO	DLATE:	524		
		(xi)	SEQU	JENCE	DESCR	IPTION	N: SEQ	ID NO:	51	
		GTTAGI	ATGA	GTGT	CGAACA	GCCTC	CAGGA	CCCCCC	CTCC	40
5		CGGGAG	AGCC	ATAG:	IGGTCT	GCGGZ	ACCGG	TGAGTA	CACC	80
		GGAATT	GCCG	GGAT	GACCGG	GTCCI	TTCTT	GGATAA	ACCC	120
		GCTCAA	TGCC	CGGA	SATTTG	GGCGT	GCCCC.	CGCGAG	ACTG	160
		CTAGCC	GAGT	AGTG	rtgggt					180
10				_						
	(2)	INFORM	ATION	FOR	SEQ II	O NO:	52			
	•	(i)	SEQU	ENCE	CHARAC	CTERIS	TICS:			
			(A)	LEN	IGTH:	טת 549	cleot	ides		
15			(B)	TY	E: nuc	cleic	acid			
			(C)	STE	RANDEDI	NESS:	sing	le		
			(D)	TOI	POLOGY	line	ar			•
		(ii)	MOLE	CULE	TYPE:	DNA				
20			•							
		(vi)	ORIG	INAL	SOURCE	E: (A	TCC #	40394)		
			(C)	INI	IVIDUA	L ISO	LATE:	hcvl		

		(xi)	SEQ	JENCE	DESCR.	IPTIO	N: SEQ	ID NO:	52	
								AACAAAC		40
								TCCCGGG		80
								GCCGCGC		120
5								AAGACTT		160
								CTATCCC		200
		GGCTCG'	TCGG	CCCG	AGGGCA	GGAC	CTGGGC	TCAGCCC	GGG	240
		TACCCT'	IGGC	CCCT	CTATGG	CAAT	GAGGGC	TGCGGGT	GGG	280
								GGCCTAG		320
10								CAATTTG		360
								GCCGACC		400
								TTGGAGG		440
٠								TCTGGAA		480
								GGTTGCT		520
15		TCTCTA:								549
	(2)	INFORM	ATIO	I FOR	SEQ II	D NO:	53			
		(i)	SEQ	JENCE	CHARAC	CTERIS	STICS:			
20			(A)	LE	NGTH:	549 ni	ucleot	ides		
			(B)	TY	PE: nu	cleic	acid			
			(C)	ST	RANDEDI	NESS:	sing	le		
-			(D)	TO	POLOGY	: line	ear			
25		(ii)	MOLI	ECULE	TYPE:	DNA				
	-	(vi)	ORIC	SINAL	SOURCE	E:				

(C)	INDIVIDUAL	ISOLATE:	us5
-----	------------	----------	-----

	(Xi) SEQ	UENCE DESCR	IPTION: SEQ	ID NO: 53	
	ATGAGCACGA	ATCCTAAACC	TCAAAGAAAA	ACCAAACGTA	4
5	ACACCAACCG	TCGCCCACAG	GACGTCAAGT	TCCCGGGTGG	8
	CGGTCAGATC	GTTGGTGGAG	TTTACTTGTT	GCCGCGCAGG	120
	GGCCCTAGAT	TGGGTGTGCG	CGCGACGAGG	AAGACTTCCG	160
	AGCGGTCGCA	ACCTCGAGGT	AGACGTCAGC	CTATCCCCAA	200
	GGCGCGTCGG	CCCGAGGGCA	GGACCTGGGC	TCAGCCCGGG	240
10	TACCCTTGGC	CCCTCTATGG	CAATGAGGGT	TGCGGGTGGG	280
	CGGGATGGCT	CCTGTCTCCC	CGTGGCTCTC	GGCCTAGTTG	320
••	GGGCCCCACA	GACCCCGGC	GTAGGTCGCG	CAATTTGGGT	360
	AAGGTCATCG	ATACCCTTAC	GTGCGGCTTC	GCCGACCACA	400
	TGGGGTACAT	ACCGCTCGTC	GGCGCCCCTC	TTGGAGGCGC	440
15	TGCCAGGGCT	CTGGCGCATG	GCGTCCGGGT	TCTGGAAGAC	480
	GGCGTGAACT	ATGCAACAGG	GAACCTTCCT	GGTTGCTCTT	520
	ருபர்கொத்துக்குக்	CCTTCTCTCCCC	CALCOACALCAL		570

(2) INFORMATION FOR SEQ ID NO: 54

20

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 549 nucleotides
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- 25 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

		(vi)	ORIG	SINAL	SOURC	E:						
			(C)	IN	DIVIDU	AL I	ISOLA	TE:	au	ısl		
.5		(xi)	SEQ	JENCE	DESCR	IPT	ION:	SEQ	ID	NO:	54	
		ATGAGC!	ACGA	ATCC:	TAAACC	TC	aaa <i>gi</i>	AAA	ACC	AAA(CGTA	40
	·	ACACCAZ										80
		CGGTCAC										120
		GGCCCTZ									_	160
10		AGCGGT										200
		GGCGCG										240
		TACCCCI										280
		CGGGAT										320
		GGGCCCI										360
15		AAGGTCA										400
		TGGGGT										440
		TGCCAG										480
		GGCGTGA										520
		TCTCTAT										549
20		10101										
20	(2)	INFORM	TIO	N FOR	SEQ II	D NO): 55	;		-		
		(i)	SEQU	JENCE	CHARA	CTE	RISTI	cs:				
			(A)	LE	NGTH:	549	nucl	eoti	ides	3		
25			(B)	TYI	E: nu	clei	ic ac	id				
			(0)	CTT	וחשתמבי	MESS	2: 5	ing	e			

		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
5		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: sp2	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55	
		ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA	4
		ACACCAACCG TCGCCCACAG GACGTCAAGT TCCCGGGTGG	8
10		CGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG	12
		GGCCCTAGAT TGGGTGTGCG CACGACGAGG AAGACTTCCG	16
		AGCGGTCGCA ACCTCGAGGT AGACGTCAGC CCATCCCCAA	20
		GGCTCGTCGA CCCGAGGGCA GGACCTGGGC TCAGCCCGGG	24
-		TACCCTTGGC CCCTCTATGG CAATGAGGGC TGCGGGTGGG	28
15		CGGGATGGCT CCTGTCTCCC CGTGGCTCTC GGCCTAGCTG	32
		GGGCCCCACA GACCCCCGGC GTAGGTCGCG CAATTTGGGT	360
		AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA	40(
		TGGGGTACAT ACCGCTCGTC GGCGCCCCTC TTGGAGGCGC	44(
		TGCCAGAGCC CTGGCGCATG GCGTCCGGGT TCTGGAAGAC	480
20		GGCGTGAACT ATGCAACAGG GAACCTTCCC GGTTGCTCTT	520
		TCTCTATCTT CCTTCTGGCC CTGCTCTCT	549
	(2)	INFORMATION FOR SEQ ID NO: 56	
25		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 549 nucleotides	
		(B) TYPE: nucleic acid	

			(C)	STR	ANDEDN	TESS:	singl	le		
			(D)	TOP	OLOGY:	linea	ar			
5		(ii)	MOLEC	ULE !	TYPE:	DNA				
3		(vi)	ORIGI		SOURCE IVIDUA	: L ISOI	LATE:	gm2	, abo	
	. 1 (140	(xi)						ID NO:	56	
10		ATGAGC		TCCT	AAACC	TCAAAC	BAAGA	ACCAAA	CGTA	4 (
		ACACCA	ACCG I	CGCC	CACAG	GACGTO	CAAGT	TCCCGG	GTGG	80
		CGGTCA	GATC G	TTGG:	IGGAG	TTTACT	TGTT	GCCGCG	CAGG	120
		GGCCCT	AGAT T	GGGT	GTGCG	CGCGAC	CGAGG	AAGACT'	TCCG	160
		AGCGGT	CGCA A	CCTC	GAGGT	AGACG1	CAGC	CTATCC	CCAA	200
15								TCAGCC		240
								TGCGGG		280
								GGCCTA		320
								CAATTT		360
								GCCGAC		400
20								TTGGAG		44(
								TCTGGA		480
								GGTTGC		520
		TCTCTA	- 1							549
25	(2)	INFORM	ATION	FOR a	SEQ II	NO:	57			
		(i)	SEQUE	NCE (CHARAC	CTERIST	TICS:			

- 101 -

		(A) LENGTH: 549 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
5			
		(ii) MOLECULE TYPE: DNA	
		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: i21	
10		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57	
		ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA	40
		ACACCAACCG TCGCCCACAG GACGTCAAGT TCCCGGGTGG	80
		CGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG	120
		GGCCCTAGAT TGGGTGTGCG CGCGACGAGG AAGACTTCCG	160
15	•	AGCGGTCGCA ACCTCGTGGT AGACGCCAGC CTATCCCCAA	200
		GGCGCGTCGG CCCGAGGGCA GGACCTGGGC TCAGCCCGGG	240
		TACCCTTGGC CCCTCTATGG CAATGAGGGT TGCGGGTGGG	280
		CGGGATGGCT CCTGTCTCCC CGTGGCTCTC GGCCTAGCTG	320
		GGGCCCCACA GACCCCCGGC GTAGGTCGCG CAATTTGGGT	360
20		AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA	400
		TGGGGTACAT ACCGCTCGTC GGCGCCCCTC TTGGAGGCGC	440
		TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAAGAC	480
		GGCGTGAACT ATGCAACAGG GAACCTTCCT GGTTGCTCTT	520
		TTTCTATTTT CCTTCTGGCC CTGCTCTCT	549
25		-	
	(2)	INFORMATION FOR SEQ ID NO: 58	

	(1) SEC	DENCE CHARACTERISTICS.	
	(A)	LENGTH: 549 nucleotides	
	(B)	TYPE: nucleic acid	
5	(C)	STRANDEDNESS: single	
	(D)	TOPOLOGY: linear	
	(ii) MOI	LECULE TYPE: DNA	
	(vi) ORI	GINAL SOURCE:	
10	(C)		
	(xi) SEQ	QUENCE DESCRIPTION: SEQ ID NO: 58	
	ATGAGCACGA	ATCCTAAACC TCAAAGAAAA ACCAAACGTA	4(
	ACACCAACCG	CCGCCCACAG GACGTTAAGT TCCCGGGCGG	80
15	TGGCCAGGTC	GTTGGTGGAG TTTACCTGTT GCCGCGCAGG	120
	GGCCCCAGGI	TGGGTGTGCG CGCGACTAGG AAGACTTCCG	160
	AGCGGTCGCA	ACCTCGTGGA AGGCGACAAC CTATCCCCAA 2	200
,	GGCTCGCCAG	CCCGAGGGCA GGGCCTGGGC TCAGCCCGGG 2	24(
	TACCCTTGGC	CCCTCTATGG CAATGAGGGT ATGGGGTGGG 2	280
20	CAGGATGGCT	CCTGTCACCC CGTGGCTCTC GGCCTAGTTG	320
	GGGCCCCACG	GACCCCGGC GTAGGTCGCG TAATTTGGGT	360
	AAGGTCATCG	ATACCCTCAC ATGCGGCTTC GCCGACCTCA	100
	TGGGGTACAT	TCCGCTCGTC GGCGCCCCCC TTAGGGGCGC 4	14(
	TGCCAGGGCC	TTGGCGCATG GCGTCCGGGT TCTGGAGGAC	180
25	GGCGTGAACT	ACGCAACAGG GAATCTGCCC GGTTGCTCCT	520
	መመመረጥን ጥረጥጥ	CCTCTTGCCT CTGCTGTCC	549

520

- 103 -

	(2)	INFORM	MATION	FOR SEQ	ID NO: 59		
		(i)	SEQUE	ince chari	ACTERISTICS	:	
		\ -•	-		549 nucleo		
5			(B)	TYPE: ni	cleic acid		
			• •		ONESS: sin		
				TOPOLOGY		3	
			(0)				
		(ii)	MOLEC	ULE TYPE:	DNA		
10		(,					
••		(vi)	ORIGI	NAL SOUR	ਾ ਸ :		
		(• •)			JAL ISOLATE	· ihl	
			(0)			•	
		(xi)	SEOUE	NCE DESCE	RIPTION: SE	Q ID NO: 59	
15		•	_			A ACCAAACGTA	4(
						r TCCCGGGCGG	_
						r gccgcgcagg	
						G AAGACTTCCG	
			* ***			C CTATCCCCAA	
20			• .			C TCAGCCCGGG	240
						r Atggggtggg	280
						C GGCCTAGTTG	320
						TAATTTGGGT	
						C GCCGACCTCA	
25						TAGGGGGGCGC	
		morran	mon r	אריז מיזיניייי	· r::00m000000	יין אריזייז אריבייטוריניטו וו	,, 47

GGCGTGAACT ATGCAACAGG GAATTTGCCC GGTTGCTCTT

		TCTCTATCTT CCTCTTGGCT CTGCTGTCC		549
	(2)	INFORMATION FOR SEQ ID NO: 60		
5		(i) SEQUENCE CHARACTERISTICS:		
		(A) LENGTH: 549 nucleotid	eś	
		(B) TYPE: nucleic acid		
	*	(C) STRANDEDNESS: single		
		(D) TOPOLOGY: linear		
10				
		(ii) MOLECULE TYPE: DNA		
		(vi) ORIGINAL SOURCE:		
		(C) INDIVIDUAL ISOLATE:	nac5	
15				
		(xi) SEQUENCE DESCRIPTION: SEQ I		
		ATGAGCACAA ATCCTAAACC CCAAAGAAAA A		40
		ACACCAACCG TCGCCCACAG GACGTCAAGT TO		80
		TGGTCAGATC GTTGGTGGAG TTTACCTGTT G	CCGCGCAGG	120
20		GGCCCCAGGT TGGGTGTGCG CGCGACTAGG A	AGACTTCCG	160
		AGCGGTCGCA ACCTCGTGGA AGGCGACAAC C	PATCCCCAA	200
		GGCTCGCCGG CCCGAGGGCA GGTCCTGGGC TO	CAGCCCGGG	240
		TACCCTTGGC CCCTCTATGG CAACGAGGGT AT	rggggtggg	280
		CAGGATGGCT CCTGTCACCC CGCGGCTCCC GC	CCTAGTTG	320
25		GGGCCCCACG GACCCCCGGC GTAGGTCGCG TA	\ATTTGGGT	360
		AAGGTCATCG ATACCCTCAC ATGCGGCTTC GO	CGACCTCA	400

- 105 -

	•	TGGGGTACAT TCCGCTCGTC GGCGCCCCCC TAGGGGGCGC	44
		TGCCAGGGCC CTGGCACATG GTGTCCGGGT TCTGGAGGAC	480
		GGCGTGAACT ATGCAACAGG GAATTTGCCT GGTTGCTCTT	520
		TCTCTATCTT CCTCTTGGCT CTGCTGTCC	549
5			
	(2)	INFORMATION FOR SEQ ID NO: 61	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 549 nucleotides	
		(B) TYPE: nucleic acid	
10		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		• • • • • • • • • • • • • • • • • • • •	
		(ii) MOLECULE TYPE: DNA	
15		(vi) ORIGINAL SOURCE:	
*, ****		(C) INDIVIDUAL ISOLATE: arg2	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61	
			40
20		ACACCAACCG CCGCCCACAG GACGTCAAGT TCCCGGGCGG	
		TGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG	
		GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG	160
		AGCGGTCGCA ACCTCGTGGA AGGCGACAAC CTATCCCCAA	200
		GGCTCGCCAG CCCGAGGGTA GGGCCTGGGC TCAGCCCGGG	240
25		TACCCTTGGC CCCTCTATGG CAATGAGGGT ATGGGGTGGG	280
		CACCOMOCOM COMOMOCOMOCO COCOMACOMO	220

		GGGCCCCACA GACCCCCGGC GTAGGTCGCG TAATTIGGGI	500
		AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA	400
		TGGGGTACAT TCCGCTCGTC GGCGCCCCCC TAGGGGGCGC	440
		TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAGGAC	480
5		GGCGTGAACT ATGCAACAGG GAATCTGCCC GGTTGCTCTT	520
J		TCTCTATCTT CCTCTTGGCT TTGCTGTCC	549
	(2)		
	(-/		
		(i) SEQUENCE CHARACTERISTICS:	
10		(A) LENGTH: 549 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
15		(ii) MOLECULE TYPE: DNA	
		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: spl	
20		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62	
		ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA	40
		ACACCAACCG CCGCCCACAG GACGTCAAGT TCCCGGGCGG	80
		TGGTCAGATC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG	120
		GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG	160
2 5		AGCGGTCGCA ACCTCGTGGA AGGCGACAAC CTATCCCCAA	200
20			240
		TATELON TO CONTRACT CARTES CONTRACT CTGGGGTGGG	280

		CAGGA	TGGCT C	CTGTCACC	c ceces	CTCTC	GGCCTAGCTG	320
		GGGCC	CTACC 6	SACCCCCGG	C GTAGG	TCGCG	CAACTTGGGT	360
	•	AAGGT	CATCG A	TACCCTTA	C GTGCG	GCTTC	GCCGACCTCA	400
		TGGGG	TACAT T	CCGCTCGT	GGCGC(cccc	TTAGGGGCGC	440
5		TGCCA	GGGCC C	TGGCGCAT	GCGTC	CGGGT	TCTGGAGGAC	480
		GGCGT	GAACT A	TGCAACAGO	GAATT	IGCCC	GGTTGCTCTT	520
		TCTCT	ATCTT C	CTCTTGGC	TTGCT	STCC		549
					£.,,			
	(2)	INFOR	MATION	FOR SEQ 1	D NO:	53		
10				,				
		(i)	SEQUE	NCE CHAR	CTERIS	CICS:		
		٠.	(A)	LENGTH:	549 nuc	leoti	ides	
			(B)	TYPE: nu	cleic a	cid		
	r		(C)	STRANDEL	NESS:	singl	Le	
15			(D)	TOPOLOGY	: linea	ır		
							•	
		(ii)	MOLEC	ULE TYPE:	DNA			
•								
		(vi)	ORIGI	NAL SOURC	E:			
20			(C)	INDIVIDU	AL ISOL	ATE:	ghl	
	••							**
		(xi)	SEQUE	NCE DESCR	IPTION:	SEQ	ID NO: 63	
		ATGAGO	CACGA A	TCCTAAACC	TCAAAG	AAAA	ACCAAACGTA	40
		ACACCA	lacce c	CGCCCACAG	GACGTO	AAGT	TCCCGGGCGG	80
25		TGGTCA	GATC G	TTGGTGGAG	TTTACT	TGTT	GCCGCGCAGG	120
		GGCCCC	AGGT T	GGTGTGCG	CGCGAC	TAGG	AAGACTTCCG	160
		N CCCCT	CCCA A		AGGCGA	CARC	CONTROCCO N	200

		GGCTCGCCGG CCCGAGGGCA GGGCCTGGGC TCAGCCCGGG	240
		TACCCTTGGC CCCTCTATGG CAATGAGGGT ATGGGGTGGG	280
		CAGGATGGCT CCTGTCACCC CGTGGTTCTC GGCCTAGTTG	320
		GGGCCCCACG GACCCCCGGC GTAGGTCGCG CAATTTGGGT	360
5		AAGATCATCG ATACCCTCAC GTGCGGCTTC GCCGACCTCA	400
		TGGGGTACAT TCCGCTCGTC GGCGCCCCC TAGGGGGCGC	440
	÷	TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAGGAC	480
		GGCGTGAACT ATGCAACAGG GAATCTGCCC GGTTGCTCCT	520
		TTTCTATCTT CCTTCTGGCT TTGCTGTCC	549
10			
	(2)	INFORMATION FOR SEQ ID NO: 64	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 549 nucleotides	
15		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
20		· · · · · · · · · · · · · · · · · · ·	
		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: 115	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64	
25		ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA	40
		ACACCAACCG CCGCCCACAG GACGTCAAGT TCCCGGGCGG	80
		TGGTCAGATC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG	120

		GGCC	CCAGGT	TGGGTGTGCG	CGCGACTAGG	AAGACTTCCG	160
		AGCGG	STCGCA	ACCTCGTGGA	AGGCGACAAC	CTATCCCCAA	200
		GGCT	CGCCAG	CCCGAGGGCA	GGGCCTGGGC	TCAGCCCGGG	240
		TACCO	CTGGC	CCCTCTATGG	CAATGAGGGI	ATGGGGTGGG	280
5		CAGG	TGGCT	CCTGTCACCC	CGCGGCTCCC	GCCTAGTTG	320
		GGGCC	CCAAA	GACCCCGGC	GTAGGTCGCG	TAATTTGGGT	360
•		AAGGT	CATCG	ATACCCTCAC	ATGCGGCTTC	GCCGACCTCA	400
		TGGGG	TACAT	TCCGCTCGTC	GGCGCCCCT	TAGGGGGCGC	440
		TGCCA	GGGCC	CTGGCGCATG	GCGTCCGGGT	TCTGGAGGAC	480
10		GGCGI	GAACT	ATGCAACAGG	GAATCTACCC	GGTTGCTCTT	520
		TCTCI	ATCTT	CCTCTTGGCT	TTGCTGTCC		549
	(2)	INFOR	MATION	FOR SEQ I	D NO: 65		
15		(i)	SEQU	ENCE CHARA	CTERISTICS:		
			(A)	LENGTH:	549 nucleot	ides	
			(B)	TYPE: nuc	cleic acid		
			(C)	STRANDEDI	NESS: sing	le	
			(D)	TOPOLOGY	linear		
20							
		(ii)	MOLE	CULE TYPE:	DNA		
		(vi)	ORIG	INAL SOURCE	:		
	٠		(C)	INDIVIDUA	L ISOLATE:	i10	-
25	•						
		(xi)	SEQU.	ENCE DESCRI	PTION: SEQ	ID NO: 65	
		ATCACO	מסמי	ስጥ ር ርጥል ስ አርር	MC73373333	7007777077	40

- 110 -

		ACACTA	ACCG	CCGCC	CACAG	GACG	TCAAGT	TCCCGGGCGG	80
		TGGCCA	GATC	GTTGG	CGGAG	TATA	CTTGCT	GCCGCGCAGG	120
		GGCCCG	AGAT	TGGGI	GTGCG	CGCG	ACGAGG	AAAACTTCCG	160
		AACGAT	CCCA	GCCAC	GCGGA	AGGC	GTCAGC	CCATCCCTAA	200
5		AGATCG'	TCGC	ACCGC	TGGCA	AGTC	CTGGGG	AAGGCCAGGA	240
J		TATCCT	TGGC	CCCTG	TATGG	GAAT	GAGGGT	CTCGGCTGGG	280
		CAGGGT	GGCT	CCTGI	CCCCC	CGTG	GCTCTC	GCCCTTCATG	320
		GGGCCC	CACT	GACCO	CCGGC	ATAG	ATCGCG	CAACTTGGGT	360
		AAGGTC	ATCG	ATACC	CTAAC	GTGC	GGTTTT	GCCGACCTCA	400
10		TGGGGT	ACAT	TCCCG	TCATC	GGCG	CCCCG	TTGGAGGCGT	440
10		TGCCAG	AGCT	CTCGC	CCACG	GAGT	GAGGGT	TCTGGAGGAT	480
		GGGGTA	AATT	ATGCA	ACAGG	GAAT	TTGCCC	GGTTGCTCTT	520
		TCTCTA							549
		101011							
15	(2)	INFORM	ATIO	FOR	SEQ I	D NO:	66		
		(i)	SEQU				STICS:		
			(A)				ucleot	ides	
					E: nu				
20			(C)	ST	RANDED	NESS:	sing	le	
			(D)	TOI	OLOGY	: lin	ear		
		(ii)	MOL	ECULE	TYPE:	DNA	1		
25		(vi)	ORI	GINAL	SOURC	E:			
		• •					OLATE:	arg6	

		(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 66	
		ATGAGCACAA ATCCTCAACC TCAAAGAAAA ACCAAAAGAA	40
		ACACTAACCG CCGCCCACAG GACGTCAAGT TCCCGGGCGG	80
		TGGTCAGATC GTTGGCGGAG TATACTTGTT GCCGCGCAGG	120
5		GGCCCCAGGT TGGGTGTGCG CGCGACGAGG AAAACTTCCG	160
		AACGGTCCCA GCCACGTGGG AGGCGCCAGC CCATCCCCAA	200
		AGATCGGCGC ACCACTGGCA AGTCCTGGGG GAAGCCAGGA	240
		TACCCTTGGC CCCTGTATGG GAATGAGGGT CTCGGCTGGG	280
		CAGGGTGGCT CCTGTCCCCC CGCGGTTCTC GCCCTTCATG	320
10 .		GGGCCCCACT GACCCCCGGC ATAGATCACG CAACTTGGGT	360
		AAGGTCATCG ATACCCTAAC GTGTGGTTTT GCCGACCTCA	400
		TGGGGTACAT TCCCGTCGGT GGTGCCCCCG TTGGTGGTGT	440
		CGCCAGAGCC CTTGCCCATG GGGTGAGGGT TCTGGAAGAC	480
		GGGATAAATT ATGCAACAGG GAATCTGCCC	510
15			
	(2)	INFORMATION FOR SEQ ID NO: 67	
		·	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 29 nucleotides	
20		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(44) MOLEGINE WIDE. DVA	
25		(ii) MOLECULE TYPE: DNA	
25		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67	
		CANACOMARC ACCIACCODO COCOACACO	20

	(2)	INFORMATION FOR SEQ ID NO: 68	
		(i) SEQUENCE CHARACTERISTICS:	
5		(A) LENGTH: 24 nucleotides	
•		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
10		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68	
		ACAGAYCCGC AKAGRTCCCC CACG	24
15	(2)	INFORMATION FOR SEQ ID NO: 69	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 30 nucleotides	
		(B) TYPE: nucleic acid	
20		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
25		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69	
		CGAACCTCGA GGTAGACGTC AGCCTATECC	30

	(2)	INFOR	MATION FOR SEQ ID NO: 70	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 30 nucleotides	
5			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
10	-	(ii)	MOLECULE TYPE: DNA	
10		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 70	
		GCAACO	CTCGT GGAAGGCGAC AACCTATCCC	30
	(2)	INFORM	MATION FOR SEQ ID NO: 71	
15				
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 30 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
20			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 71	
25		GTCACC	AATG ATTGCCCTAA CTCGAGTATT	30
	(2)	INFORM	ATION FOR SEO ID NO: 72	

		(i)	SEQUENCE CHARACTERISTICS:	
		٠	(A) LENGTH: 26 nucleotides	
			(B) TYPE: nucleic acid	
5			(C) STRANDEDNESS: single	
-			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
10	•	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 72	
		GTCAC	GAACG ACTGCTCCAA CTCAAG	26
	(2)	INFOR	MATION FOR SEQ ID NO: 73	
15	-	(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 28 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
•			(D) TOPOLOGY: linear	
20				
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 73	
			ATGAT CGCTGGWGCY CACTGGGG	28
25				
	(2)	TATEODI	MATTON FOR SEO ID NO: 74	

		(i)	SEQUENCE CHARACTERISTICS:	
		•	(A) LENGTH: 28 nucleotides	
			(B) TYPE: nucleic acid	
5			(C) STRANDEDNESS: single	
•			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 74	
10		TGGAY	ATGGT GGYGGGGCY CACTGGGG	28
	(2)	INFORM	MATION FOR SEQ ID NO: 75	
		(i)	SEQUENCE CHARACTERISTICS:	
15			(A) LENGTH: 20 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
20		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 75	
		ATGATG	AACT GGTCVCCYAC	20
2 5	(2)	INFORM	ATION FOR SEQ ID NO: 76	
		(i)	SEOUENCE CHARACTERISTICS:	

			(A) LENGTH: 26 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
5				
		(ii)	MOLECULE TYPE: DNA	
			SEQUENCE DESCRIPTION: SEQ ID NO: 76	
			GCCC AGTTSCCCRC CATGGA	26
10	(2)	INFORM	ATION FOR SEQ ID NO: 77	
	•-•			
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 22 nucleotides	
			(B) TYPE: nucleic acid	
15			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
20		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 77	
		AACCCA	CTCT ATGYCCGGYC AT	22
	(2)	INFORM	ATION FOR SEQ ID NO: 78	
25		(i)	SEQUENCE CHARACTERISTICS:	
-			(A) LENGTH: 18 nucleotides	
			(B) TYPE: nucleic acid	

			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
5		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 78	
		GAATC	SCTGG GGTGACCG	18
10	(2)	INFORM	MATION FOR SEQ ID NO: 79	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 28 nucleotides	
			(B) TYPE: nucleic acid	,
			(C) STRANDEDNESS: single	
15			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 75	
20		CCATGA	ATCA CTCCCCTGTG AGGAACTA	28
	(2)	INFORM	ATION FOR SEQ ID NO: 80	
		(i)	SEQUENCE CHARACTERISTICS:	
25			(A) LENGTH: 18 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	

		•	(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
5.	(2)	TTGCGG	SEQUENCE DESCRIPTION: SEQ ID NO: 80 GGGGC ACGCCCAA MATION FOR SEQ ID NO: 81	18
10			SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	-
15		(ii)	MOLECULE TYPE: DNA	
20	(2)	YGAAGO	SEQUENCE DESCRIPTION: SEQ ID NO: 81 CGGGC ACAGTCARRC AAGARAGCAG GGC	33
25		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single	v

		:	(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
5			SEQUENCE DESCRIPTION: SEQ ID NO: 82	
	•	RTARA	GCCCY GWGGAGTTGC GCACTTGGTR GGC	33
:	(2)	INFOR	MATION FOR SEQ ID NO: 83	
	•	(i)	SEQUENCE CHARACTERISTICS:	
10			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
15		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 83	
		RATACI	ICGAG TTAGGGCAAT CATTGGTGAC RTG	33
20	(2)	INFORM	MATION FOR SEQ ID NO: 84	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
25			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: .linear	

		(ii)	MOLECULE TYPE: DNA	
			SEQUENCE DESCRIPTION: SEQ ID NO: 84 SCAGG ATGGYATCRK BCGYCTCGTA CAC	33
5				
	(2)		MATION FOR SEQ ID NO: 85	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
•	•		(B) TYPE: nucleic acid	
10			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
15			SEQUENCE DESCRIPTION: SEQ ID NO: 85	
		GTTRCC	CCTCR CGAACGCAAG GGACRCACCC CGG	33
	(2)	INFORM	MATION FOR SEQ ID NO: 86	
20		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25				
_ •		(ii)	MOLECULE TYPE: DNA	

	•	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 86	
		CGTR	GGGTY AYCGCCACCC AACACCTCGA GRC	33
	(2)	INFOR	MATION FOR SEQ ID NO: 87	
5				
		(i)	SEQUENCE CHARACTERISTICS:	
		•	(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
		•	(C) STRANDEDNESS: single	
10			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 87	
15		CGTYG	YGGGG AGTTTGCCRT CCCTGGTGGC YAC	33
	(2)	INFOR	MATION FOR SEQ ID NO: 88	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
•			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 88	

		CCCGA	CAAGC AGATCGATGT GACGTCGAAG CTG	33
	(2)	INFOR	MATION FOR SEQ ID NO: 89	
5		(i)	SEQUENCE CHARACTERISTICS:	
	-		(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
10				
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 89	
٠		CCCCA	CGTAG ARGGCCGARC AGAGRGTGGC GCY	33
15				
٠	(2)	INFOR	MATION FOR SEQ ID NO: 90	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) / LENGTH: 33 nucleotides	
20			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 90	
			CGACA AGAAAGACAG ACCCGCAYAR GTC	. 33

	(2)	INFOR	RMATION FOR SEQ ID NO: 91	
		(i)	SEQUENCE CHARACTERISTICS:	
5			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
		•	(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
10		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 91	
		CGTCC	AGTGG YGCCTGGGAG AGAAGGTGAA CAG	33
15	(2)	INFOR	MATION FOR SEQ ID NO: 92	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
20			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
25		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 92	
		GCCGG	GATAG ATRGARCAAT TGCARYCTTG CGT	33

	(2)	INFOR	MATION FOR SEQ 1D NO: 93	
•		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
5			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
		•	(D) TOPOLOGY: linear	
7.0		(ii)	MOLECULE TYPE: DNA	
10		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 93	
	•	CATAT	CCCAT GCCATGCGGT GACCCGTTAY ATG	33
	(2)	INFOR	MATION FOR SEQ ID NO: 94	
15				
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
20			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
			SEQUENCE DESCRIPTION: SEQ ID NO: 94	
25		YACCA	AYGCC GTCGTAGGGG ACCARTTCAT CAT	33
	(2)	INFOR	MATION FOR SEQ ID NO: 95	

		(i) SEQUENCE	CHARACTERISTICS:	
		(A) LE	NGTH: 33 nucleotides	
		(B) TY	PE: nucleic acid	
5		(C) ST	RANDEDNESS: single	**
		(D) TO:	POLOGY: linear	
		(ii) MOLECULE	TYPE: DNA	
		(xi) SEQUENCE	DESCRIPTION: SEQ ID NO: 95	
10		GATGGCTTGT GGGA:	TCCGGA GYASCTGAGC YAY	33
	(2)	INFORMATION FOR	SEQ ID NO: 96	
		(i) SEQUENCE	CHARACTERISTICS:	
15		(A) LE	NGTH: 33 nucleotides	
		(B) TYI	PE: nucleic acid	
		(C) STI	RANDEDNESS: single	
		(D) TO	POLOGY: linear	
20		(ii) MOLECULE	TYPE: DNA	
		(xi) SEQUENCE	DESCRIPTION: SEQ ID NO: 96	
		GACTCCCCAG TGRGC	CWCCAG CGATCATRTC CAW	33
25	(2)	INFORMATION FOR	SEQ ID NO: 97	
		(i) SEQUENCE	CHARACTERISTICS:	

	(A) LENGTH: 33 nucleotides	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
5		
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97	
	CCCCACCATG GAGAAATACG CTATGCCCGC YAG	33
10 (2)	INFORMATION FOR SEQ ID NO: 98	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 33 nucleotides	
	(B) TYPE: nucleic acid	
15	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98	
	TAGYAGCAGY ACTACYARGA CCTTCGCCCA GTT	33
- (2)	INFORMATION FOR SEQ ID NO: 99	
25	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 33 nucleotides	
	(B) TYPE: nucleic acid	

- 127 -

			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
5		(ii)	MOLECULE TYPE: DNA	
		•	SEQUENCE DESCRIPTION: SEQ ID NO: 99 GTGR GTKTCYGCGT CRACGCCGGC RAA	33
10	(2)	INFORM	ATION FOR SEQ ID NO: 100	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
15			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 100	
20		GGAAGY	TGGG ATGGTYARRC ARGASAGCAR AGC	33
	(2)	INFORM	ATION FOR SEQ ID NO: 101	
		(i)	SEQUENCE CHARACTERISTICS:	
25			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	

			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
5	(2)	GTAYAY	SEQUENCE DESCRIPTION: SEQ ID NO: 101 FYCCG GACRCGTTGC GCACTTCRTA AGC MATION FOR SEQ ID NO: 102	33
10		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
15		(ii)	MOLECULE TYPE: DNA	
		AATRCI	SEQUENCE DESCRIPTION: SEQ ID NO: 102 TTGMG TTGGAGCART CGTTYGTGAC ATG	33
20	(2)	INFORM	MATION FOR SEQ ID NO: 103	
25		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	

		(11)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 103	
		RGYRT	GCATG ATCAYGTCCG YYGCCTCATA CAC	33
5				
	(2)	INFOR	MATION FOR SEQ ID NO: 104	
		·(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
	•		(B) TYPE: nucleic acid	
10			(C) STRANDEDNESS: single	
		·	(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
15		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 104	
		RTTGT	YYTCC CGRACGCARG GCACGCACCC RGG	3 3
•	(2)	INFORM	MATION FOR SEQ ID NO: 105	
20		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
2 5			. ·	
		(ii)	MOLECIUE TYPE: DNA	

			SEQUENCE DESCRIPTION: SEQ ID NO: 105 GRGTS AGCGCYACCC AGCARCGGGA GSW	33
		C0100		
	(2)	INFOR	MATION FOR SEQ ID NO: 106	
5		445		
		(1)	SEQUENCE CHARACTERISTICS:	
	•		(A) LENGTH: 33 nucleotides	
,			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
10			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 106	
15		YGTRG:	TGGGG AYGCTGKHRT TCCTGGCCGC VAR	33
	(2)	INFOR	MATION FOR SEQ ID NO: 107	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 107	

		CCCRA	ACGAGC AARTCGACRT GRCGTCGTAW TGT	33
	(2)	INFOR	MATION FOR SEQ ID NO: 108	
5		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	•
٠			(C) STRANDEDNESS: single	
		•	(D) TOPOLOGY: linear	
10				
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 108	
		YCCCA	CGTAC ATAGCSGAMS AGARRGYAGC CGY	33
15				
	(2)	INFOR	MATION FOR SEQ ID NO: 109	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
20			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
- -		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 109	
		CTGGG	AGAYR AGRAAAACAG ATCCGCARAG RTC	33

	(2)	INFOR	MATION FOR SEQ ID NO: 110	
		(i)	SEQUENCE CHARACTERISTICS:	
5		• •	(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
		. 	(D) TOPOLOGY: linear	
10		(ii)	MOLECULE TYPE: DNA	
			SEQUENCE DESCRIPTION: SEQ ID NO: 110 CRTGC CGGCCAGSBG AGAAGGTGAA YAG	33
		YGTCT	IRIGO COOCCAOSDO AGRAGOTORA TAC	
15	(2)	INFOR	MATION FOR SEQ ID NO: 111	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
20			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
25		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 111	
		GCCGG	GATAG AKKGAGCART TGCAKTCCTG YAC	33

	(2)	INFOR	MATION FOR SEQ ID NO: 112	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
5			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
	•	٠	(D) TOPOLOGY: linear	
10		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 112	
		CATAT	CCCAA GCCATRCGRT GGCCTGAYAC CTG	33
	(2)	INFOR	MATION FOR SEQ ID NO: 113	
15		(4)	SEQUENCE CHARACTERISTICS:	
		(1)	(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
20			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 113	
25		CACTA	RGGCT GYYGTRGGYG ACCAGTTCAT CAT	33

		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
5			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
	•	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 114	
10			CTTGT GGGATCCGGA GTAACTGCGA YAC	33
	(2)	INFORM	MATION FOR SEQ ID NO: 115	
,		(i)	SEQUENCE CHARACTERISTICS:	
15			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
20		(ii)	MOLECULE TYPE: DNA	
			SEQUENCE DESCRIPTION: SEQ ID NO: 115	33
			•	
25	(2)	INFORM	MATION FOR SEQ ID NO: 116	
		(i)	SEQUENCE CHARACTERISTICS:	

			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
5				
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 116	
		SCCCA	CCATG GAWWAGTAGG CAAGGCCCGC YAG	33
10	(2)	INFOR	MATION FOR SEQ ID NO: 117	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
15			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
20		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 117	
		GAGTA	SCATC ACAATCAADA CCTTAGCCCA GTT	33
	(2)	INFOR	MATION FOR SEQ ID NO: 118	
25		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	

			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
5			SEQUENCE DESCRIPTION: SEQ ID NO: 118 YGYRG GTRTKCCCGT CAACGCCGGC AAA	33
	(2)	INFOR	MATION FOR SEQ ID NO: 119	
10		/±\	SEQUENCE CHARACTERISTICS:	
		(1)	(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
15			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 119	
20	•		ACAGG GGAGTGATTC ATGGTGGAGT GTC	33
	(2)	INFOR	MATION FOR SEQ ID NO: 120	
	^	(i)	SEQUENCE CHARACTERISTICS:	
25			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	•
			(C) STRANDEDNESS: single	

		<i>i</i>	(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
5			SEQUENCE DESCRIPTION: SEQ ID NO: 12	
	401		TAGAC GCTTTCTGCG TGAAGACAGT AGT	33
	(2)	INFOR	MATION FOR SEQ ID NO: 121	
;	•	(i)	SEQUENCE CHARACTERISTICS:	
10			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
15		(ii)	MOLECULE TYPE: DNA	
		•	SEQUENCE DESCRIPTION: SEQ ID NO: 12	
		GCCTGG	GAGGC TGCACGRCAC TCATACTAAC GCC	33
20	(2)	INFORM	MATION FOR SEQ ID NO: 122	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
		-,	(B) TYPE: nucleic acid	
25			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	

		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122 CGCAGACCAC TATGGCTCTY CCGGGAGGGG GGG	33
5	(2)	<pre>INFORMATION FOR SEQ ID NO: 123 (i) SEQUENCE CHARACTERISTICS:</pre>	
		(ii) MOLECULE TYPE: DNA	
15		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123 TCRTCCYGGC AATTCCGGTG TACTCACCGG TTC	33
	(2)	INFORMATION FOR SEQ ID NO: 124	
20		 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
25		(ii) MOLECULE TYPE: DNA	

			SEQUENCE DESCRIPTION: SEQ ID NO: 124 GAGCG GGTTDATCCA AGAAAGGACC CGG	33
5	(2)	INFOR	MATION FOR SEQ ID NO: 125	
,		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
10	·		(D) TOPOLOGY: linear	
		(ii) ,	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 125	
15		AGCAG!	ICTYG CGGGGGCACG CCCAARTCTC CAG	33
	(2)	INFOR	MATION FOR SEQ ID NO: 126	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 126	

		ACAAG	GCCTT TCGCGACCCA ACACTACTCG GCT	33
	(2)	INFOR	MATION FOR SEQ ID NO: 127	
5		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
10	•			
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 127	
1			ACTCG CAAGCACCCT ATCAGGCAGT ACC	33
15	(2)	INFOR	MATION FOR SEQ ID NO: 128	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
20			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	

(2) INFORMATION FOR SEQ ID NO: 129 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129			(ii)	MOLECULE TYPE: DNA	
(2) INFORMATION FOR SEQ ID NO: 129 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (Xi) MOLECULE TYPE: DNA (Xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129 GTTACGTTTG KTTYTTYTTT GRGGTTTRGG AWT 33 (2) INFORMATION FOR SEQ ID NO: 130 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single			(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 128	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129 GTTACGTTTG KTTYTTYTTT GRGGTTTRGG AWT 33 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single	5		YGTGCT	CATG RIGCACGGIC TACGAGACCI CCC	33
(A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129 GTTACGTTTG KTTYTTYTTT GRGGTTTRGG AWT 33 (2) INFORMATION FOR SEQ ID NO: 130 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single		(2)	INFORM	MATION FOR SEQ ID NO: 129	
(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129 GTTACGTTTG KTTYTTYTTT GRGGTTTRGG AWT 33 (2) INFORMATION FOR SEQ ID NO: 130 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single			(i)	SEQUENCE CHARACTERISTICS:	
(C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129 GTTACGTTTG KTTYTTYTTT GRGGTTTRGG AWT 33 (2) (2) INFORMATION FOR SEQ ID NO: 130 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single	10			(A) LENGTH: 33 nucleotides	
(D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129 GTTACGTTTG KTTYTTYTTT GRGGTTTRGG AWT 33 (2) INFORMATION FOR SEQ ID NO: 130 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single				(B) TYPE: nucleic acid	
(D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129 GTTACGTTTG KTTYTTYTTT GRGGTTTRGG AWT 33 (2) INFORMATION FOR SEQ ID NO: 130 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single				(C) STRANDEDNESS: single	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129 GTTACGTTTG KTTYTTYTTT GRGGTTTRGG AWT 33 20 (2) INFORMATION FOR SEQ ID NO: 130 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single					
GTTACGTTTG KTTYTTYTTT GRGGTTTRGG AWT 20 (2) INFORMATION FOR SEQ ID NO: 130 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single	15		(ii)	MOLECULE TYPE: DNA	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single			(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 129	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single			GTTACG	TTTG KTTYTTYTTT GRGGTTTRGG AWT	33
(A) LENGTH: 33 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single	20	(2)	INFORM	ATION FOR SEQ ID NO: 130	
(B) TYPE: nucleic acid (C) STRANDEDNESS: single			(i)	SEQUENCE CHARACTERISTICS:	
(C) STRANDEDNESS: single				(A) LENGTH: 33 nucleotides	
(C) STRANDEDNESS: single				(B) TYPE: nucleic acid	
-	25				
				_	

		(ii)	MOLECULE TYPE: DNA	
			SEQUENCE DESCRIPTION: SEQ ID NO: 130 ACTTR ACGTCCTGTG GGCGRCGGTT GGT	33
5		Cooch	netti neeteete eeemeete te	
•	(2)	INFOR	MATION FOR SEQ ID NO: 131	
•		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
10			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
15		4 4 3	CROTHINGS DESCRIPTION: SEC ID NO: 131	
			SEQUENCE DESCRIPTION: SEQ ID NO: 131 AAACT CCACCRACGA TCTGRCCRCC RCC	33
		CARGIZ	AMACT CCACCRACGA TCTGRCCRCC RCC	
	(2)	INFOR	MATION FOR SEQ ID NO: 132	
20				
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
25			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	

		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 132	
		RCGCA	CACCC AAYCTRGGGC CCCTGCGCGG CAA	33
5	(2)	INFOR	MATION FOR SEQ ID NO: 133	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
10			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
"فلدن		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 133	
15		AGGTT	GCGAC CGCTCGGAAG TCTTYCTRGT CGC	33
	(2)	INFOR	MATION FOR SEQ ID NO: 134	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEO ID NO: 134	

		RCGHR	CCTTG GG	GATAGGCT GACGTCWACC TCG	33
	(2)	INFOR	MATION F	OR SEQ ID NO: 135	
5		(i)	SEQUEN	CE CHARACTERISTICS:	
			(A)	LENGTH: 33 nucleotides	
	•		(B)	TYPE: nucleic acid	
			(C)	STRANDEDNESS: single	
		•		TOPOLOGY: linear	•
10				•	
		(ii)	MOLECU	LE TYPE: DNA	
		(xi)	SEQUEN	CE DESCRIPTION: SEQ ID NO: 135	
		RCGHR	CCTTG GG	GATAGGTT GTCGCCWTCC ACG	33
15	(2)	INFOR	MATION F	OR SEQ ID NO: 136	
		(i)	SEQUEN	CÉ CHARACTERISTICS:	
			(A)	LENGTH: 33 nucleotides	
			(B) !	TYPE: nucleic acid	
20 .			(C)	STRANDEDNESS: single	
			(D)	TOPOLOGY: linear	
		(ii)	MOLECU	LE TYPE: DNA	
25		(xi)	SEQUEN	CE DESCRIPTION: SEQ ID NO: 136	
		YCCRG	CTGR GC	CCAGRYCC TRCCCTCGGR YYG	33

	(2)	1111 (1)	WHITON TON BEG ID NO. 137	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
5			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
		.•	(D) TOPOLOGY: linear	
10		(ii)	MOLECULE TYPE: DNA	
	·	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 137	
		BSHRC	CCTCR TTRCCRTAGA GGGGCCADGG RTA	33
15	(2)	INFOR	MATION FOR SEQ ID NO: 138	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
20			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 138	
25		GCCRC	GGGGW GACAGGAGCC ATCCYGCCCA CCC	33
	(2)	INFOR	MATION FOR SEQ ID NO: 139	

SUBSTITUTE SHEET

		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
5			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
10	•	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 139	
		CCGGGG	GTCY GTGGGGCCCC AYCTAGGCCG RGA	33
	(2)	INFORM	MATION FOR SEQ ID NO: 140	•
		(i)	SEQUENCE CHARACTERISTICS:	
15	•		(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
20		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 140	
		ATCGAT	GACC TTACCCAART TRCGCGACCT RCG	33
25	(2)	INFORM	NATION FOR SEQ ID NO: 141	
		(i)	SEQUENCE CHARACTERISTICS:	

			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
5				
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 141	
		CCCCA	TGAGR TCGGCGAAGC CGCAYGTRAG GGT	33
10				
	(2)	INFOR	MATION FOR SEQ ID NO: 142	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
•			(B) TYPE: nucleic acid	
15			(C) STRANDEDNESS: single	•
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
20		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 142	
=		GCCYC	CWARR GGGGCGCCGA CGAGCGGWAT RTA	33
	(2)	INFORM	MATION FOR SEQ ID NO: 143	
25		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	

			(C) STRANDEDNESS: single	
		•	(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
5		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 143	
		AACCC	GGACR CCRTGYGCCA RGGCCCTGGC AGC	33
	(2)	INFOR	MATION FOR SEQ ID NO: 144	
LO		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
15				
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 144	
		RTTCCC	CTGTT GCATAGTTCA CGCCGTCYTC CAG	33
20	(2)	INFORM	MATION FOR SEQ ID NO: 145	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
:5			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	

		(ii)	MOLECULE TYPE: DNA	
5			SEQUENCE DESCRIPTION: SEQ ID NO: 145	33
	(2)	INFOR	MATION FOR SEQ ID NO: 146	
		(i)	SEQUENCE CHARACTERISTICS:	
10			(A) LENGTH: 20 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
15		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 146	
		AGGCA	TAGGA CCCGTGTCTT	20
20	(2)	INFOR	MATION FOR SEQ ID NO: 147	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 20 nucleotides	
			(B) TYPE: nucleic acid	
25			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 147	
		CTTCT	TTGGA GAAAGTGGTG	20

SUBSTITUTE SHEET

CLAIMS

- 1. As a composition of matter, a non-naturally occurring nucleic acid having a non-HCV-1 nucleotide sequence of eight or more nucleotides corresponding to a nucleotide sequence within the hepatitis C virus genome.
- 2. The composition of claim 1 wherein said nucleotide sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome is selected from the regions consisting of the NS5 region, envelope 1 region, 5'UT region, and the core region.
- 3. The composition of claim I wherein said nucleotide sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the NS5 region.
- 20 4. The composition of claim 3 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome is selected from a sequence within sequences numbered 2-22.

5. The composition of claim 1 wherein said nucleotide sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the envelope 1 region.

5

6. The composition of claim 5 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome corresponds to a sequence within sequence numbers 24-32.

10

7. The composition of claim 1 wherein at least one sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the 5'UT region.

15

8. The composition of claim 7 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome corresponds to a sequence within sequences numbered 34-51.

20

9. The composition of claim 1 wherein said nucleotide sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the core region.

10. The composition of claim 9 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome corresponds to a within sequences numbered 53-66.

5

11. The composition of claim 1 wherein said non-naturally occurring nucleic acid has a nucleotide sequence corresponding to one or more genotypes of hepatitis C virus.

10

- 12. The composition of claim 11 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, 33-38 in the 5'UT region, and 52-57 in the core region.
- 13. The composition of claim 11 wherein said
 20 non-naturally occurring nucleic acid has a sequence
 corresponding to a sequence of a second genotype which
 second genotype is defined substantially by sequences
 numbered 7-12 in the NS5 region, 26-28 in the envelope
 1 region, 39-45 in the 5'UT region, and 58-64 in the
 25 core region.

- 14. The composition of claim 11 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, 46-47 in the 5'UT region and 65-66 in the core region.
- 15. The composition of claim 11 wherein said
 10 non-naturally occurring nucleic acid has a sequence
 corresponding to a sequence of a fourth genotype which
 fourth genotype is defined substantially by sequences
 numbered 20-22 in the NS5 region, 29-31 in the envelope
 1 region and 48-49 in the 5'UT region.
- 16. The composition of claim 11 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the NS5 region and 50-51 in the 5'UT region.
- 17. The composition of claim'l wherein said non-naturally occurring nucleic acid is capable of
 25 priming a reaction for the synthesis of nucleic acid to form a nucleic acid having a nucleotide sequence corresponding to hepatitis C virus.

- 18. The composition of claim 1 wherein said non-naturally occurring nucleic acid has label means for detecting a hybridization product.
- 5 19. The composition of claim 1 wherein said non-naturally occurring nucleic acid has support means for separating a hybridization product from solution.
- 20. The composition of claim 1 wherein said non-naturally occurring nucleic acid prevents the transcription or translation of viral nucleic acid.
- 21. A method of forming a hybridization product with a hepatitis C virus nucleic acid comprising the following steps:
 - a. placing a non-naturally occurring nucleic acid having a nucleotide sequence of eight or more nucleotides corresponding to a non-HCV-1 sequence in the hepatitis C viral genome into conditions in which hybridization conditions can be imposed said non-naturally occurring nucleic acid capable of forming a hybridization product with said hepatitis C virus nucleic acid under hybridization conditions; and

- b. imposing hybridization conditions to form a hybridization product in the presence of hepatitis C virus nucleic acid.
- 5 22. The method of claim 21 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence in the hepatitis C virus genome corresponds to a sequence within at least one of the regions consisting essentially of NS5 region, envelope 1 region, 5'UT
- 10 region, and the core region.
 - 23. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponds to a sequence within the NS5 region.

24. The method of claim 23 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponds to a sequence within sequences numbered 2-22.

20

25. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponds to a sequence within the envelope 1 region.

- 26. The method of claim 25 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence is selected from a sequence within sequences numbered 24-32.
- The method of claim 21 wherein said nucleotide
 sequence corresponds to a non-HCV-1 sequence
 corresponding to a sequence within the 5'UT region.
- 10 28. The method of claim 27 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence selected from a sequence within sequences numbered 34-51.
- 29. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponding to a sequence within the core region.
- 30. The method of claim 29 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence selected from a sequence within sequences numbered 53-66.
 - 31. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 nucleotide sequence corresponding to one or more genotypes of hepatitis C virus.

- 32. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, 33-38 in the 5'UT region, and 52-57 in the core region.
- 33. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a second genotype which second genotype is defined substantially by sequences numbered 7-12 in the NS5 region, 26-28 in the envelope 1 region, 39-45 in the 5'UT region, and 58-64 in the core region.
- 15 34. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, 46-47 in the 5'UT region and 65-66 in the core region.
- 35. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fourth genotype which fourth genotype
 25 is defined substantially by sequences numbered 20-22 in the NS5 region, 29-31 in the envelope 1 region and 48-49 in the 5'UT region.

- 36. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the NS5 region and 50-51 in the 5'UT region.
- 37. The method of claim 21 wherein said hybridization product is capable of priming a reaction for the synthesis of nucleic acid.
- 38. The method of claim 21 wherein said non-naturally occurring nucleic acid has label means for detecting a hybridization product.
- 15 39. The method of claim 21 wherein said non-naturally occurring nucleic acid has support means for separating the hybridization product from solution.
- 40. The method of claim 21 wherein said non-naturally occurring nucleic acid prevents the transcription or translation of viral nucleic acid.
- 41. As a composition of matter, a non-naturally occurring polypeptide corresponding to a non-HCV-1 nucleotide sequence of nine or more nucleotides which sequence of nine or more nucleotides corresponds to a sequence within hepatitis C virus genomic sequences.

- 42. The composition of claim 41 wherein said non-HCV-1 sequence is selected from one of the regions consisting of NS5 region, envelope 1 region, and the core region.
- 5 43. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence corresponds to a sequence in the NS5 region.
- 44. The composition of claim 43 wherein said non-HCV-1 sequence is selected from a sequence within sequences numbered 2-22.
- 45. The composition of claim 41 wherein said non-HCV-1 sequence corresponds to a sequence in the envelope 1 region.
 - 46. The composition of claim 45 wherein said non-HCV-1 sequence is selected from a sequence within sequences numbered 24-32.

- 47. The composition of claim 41 wherein said non-HCV-1 sequence corresponds to a sequence in the core region.
- 48. The composition of claim 47 wherein said non-HCV-1 sequence is selected from a sequence within sequences numbered 52-66.

- 49. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a nucleotide sequence corresponding to one or more genotypes of hepatitis C virus.
- 50. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, and 52-57 in the core region.
- 51. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a second genotype which second genotype is defined substantially by sequences numbered 7-12 in the NS5 region, 26-28 in the envelope 1 region, and 58-64 in the core region.
- 20 52. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, and 65-66 in the core region.

. .

20

- 53. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a fourth genotype which fourth genotype is defined substantially by sequences numbered 20-22 in the NS5 region, 29-31 in the envelope 1 region and 48-49 in the 5'UT region.
- 54. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the NS5 region and 50-51 in the 5'UT region.
- 55. The composition of claim 41 wherein said
 15 polypeptide is capable of generating an immune reaction in a host.
 - 56. An antibody capable of selectively binding to the composition of claim 41.
 - 57. A method of detecting one or more genotypes of hepatitis C virus comprising the following steps:
 - a) placing a non-naturally occurring nucleic acid having a nucleotide sequence of eight or more nucleotides corresponding to one or more genotypes of hepatitis C virus under conditions where hybridization conditions can be imposed,

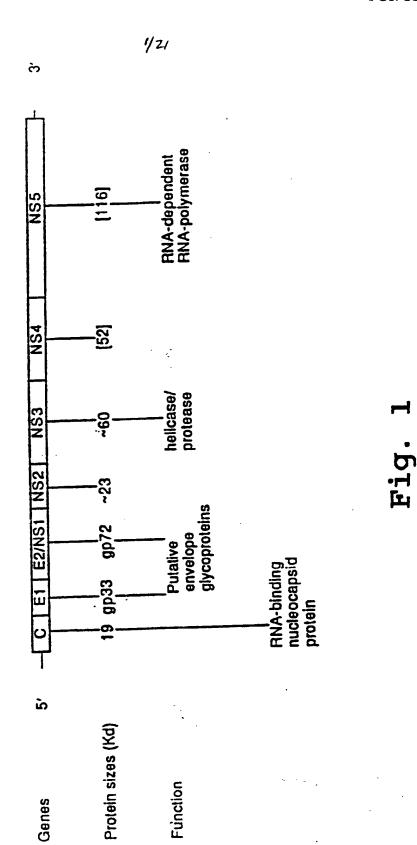
- b) imposing hybridization conditions to form a hybridization product in the presence of hepatitis
 C virus nucleic acid; and
- c) monitoring the non-naturally occurring nucleic acid for the formation of a hybridization product, which hybridization product is indicative of the presence of the genotype of hepatitis C virus.
- 58. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, 33-38 in the 5'UT region, and 52-57 in the core region.
- 59. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a second genotype which second genotype is defined substantially by sequences numbered 7-12 in the NS5 region, 26-28 in the envelope 1 region, 39-45 in the 5'UT region, and 58-64 in the core region.

- 60. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, 46-47 in the 5'UT region and 65-66 in the core region.
- 61. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fourth genotype which fourth genotype is defined substantially by sequences numbered 20-22 in the NS5 region, 29-31 in the envelope 1 region and 48-49 in the 5'UT region.
- 15 62. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the NS5 region.

20

63. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence numbered 67-145.

- 64. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence numbered 69, 71, 73 and 81-99 to identify Group I genotypes in the core and region of the HCV genome.
- 65. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence numbered 70, 72, 70 and 100-118 to identify Group II genotypes in the core and envelope regions of the HCV genome.
- 66. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to
 15 a sequence numbered 77 to identify Group III genotypes in the 5' UT region of the HCV genome.
- 67. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence numbered 79 to identify Group IV genotypes in the 5' UT region of the HCV genome.



SUBSTITUTE SHEET

2/2-1

Fig. 2a

SEQUENCE ID NUMBER	GENOTYPE	21 25 51 13 13 13	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	1	8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	CTCCACAGTC ACTGAGAGCG ACATCCGTAC GGAGGAGCA ATCTACCAAT GTTGTGACCT CGACCCCAA CTCCACAGTC ACTGAGAGCG ACATCCGTAC GGAGGAGCA ATCTACCAAT GTTGTGACCT CGACCCCCAA CTCCACAGTC ACTGAGAGCG ACATCCGTAC GGAGGAGCA ATCTACCAAT GTTGTGACT GGACCCCCAA CTCTACAGTC ACTGAGAACG ACATCCGTAC GGAGGAGCA ATTTACCAAT GTTGTGACT GGACCCCCAA CTCTACAGTC ACTGAGAACG ACATCCGTAC GGAGGAGCA ATTTACCAAT GTTGTGACCT GGACCCCCAA CTCCACAGTC ACTGAGAGCG ATATCCGTAC GGAGGAGCCA ATCTACCAGT GTTGTGACCT GGACCCCCAA CTCTACAGTC ACTGAGAGCG ATATCCGTAC GGAGGAGCCA ATCTACCAAT GTTGTGACCT GGACCCCCAA
7 7 8 9 10 11 12		8 8 8 AAAAAA	CTCCACAGTC ACTGAGAATG ACACCCGTGT TGAGGAGTCA ATTTACCAAT GTTGTGACTT CTCCACGGTC ACTGAGAATG ACACCCGTGT TGAGGAGTCA ATTTACCAAT GTTGTGACTT CTCAACGGTC ACTGAGAATG ACATCCGTGT TGAGGAGTCA ATTTACCAAA GTTGTGACTT CTCAACGGTC ACCGAGAATG ACATCCGTGT TGAGGAGTCA ATTTATCAAT GTTGTGACTT CTCAACGGTC ACTGAGAGTG ACATCCGTGT CGAGGAGTCG ATTTACCAAT GTTGTGACTT CTCCACAGTC ACTGAGAGTG ACATCCGTGT TGAGGAGTCA ATTTACCAAT GTTGTGACTT CTCCAACAGTC ACTGAGAGTG ACATCCGTGT TGAGGAGTCA ATTTACCAAT GTTGTGACTT
13 13 14 15 16	6111	;; ;; ;; ;; ;; ;; ;; ;; ;; ;;	CTCAACCGTC ACTGAGAGA ACATCAGAAC TGAGGAGTCC ATATACCGAG CCTGCTCCCT GCCTGAGGAG CTCTACAGTC ACGTAAAAGG ACATCACATC CTAGGAGTCC ATCTACCAGT CCTGTTCACT GCCCGAGGAG CTCTACAGTC ACAGAGAGG ACATCAGAAC CGAGGAGTCC ATCTATCTGT CCTGCTCACT GCCTGAGGAG CTCTACAGTC ACGGAGAGGG ACATCAGAAC CGAGGAGTCC ATCTATCTGT CCTGTTCACT GCCTGAGGAG CTCTACAGTC ACGGAGAGGG ACATCAGAAC AGAAGAATCC ATATATCAGG GTTGTTCCCT GCCTCAGGAG
18 19	ΛĐ		CTCGACCGIT ACCGAACAIG ACAIAAIGAC IGAAGAGICI AITIACCAAI CAIIGIACII G CICGACCGII ACCGAACAIG ACAIAAIGAC IGAAGAGICC AITIACCAAI CAIIGIACII G
20 21 22	010	i t	CTCTACTGTC ACTGAACAGG ACATCAGGGT GGAAGAGGAG ATATACCAGT GCTGTAACCT TGAACCGGAG CTCGACTGTC ACTGAACAGG ACATCAGGGT GGAAGAGGAG ATATACCAAT GCTGTAACCT TGAACCGGAG CTCAACTGTC ACTGAACAGG ACATCAGGGT GGAAGAGGAG ATATACCAAT GCTGTAACCT TGAACCGGAG

SUBSTITUTE SHEET

Fig. 2k

NS5 REGION - (2/5)

10 NUMBER GENOTYPE 1 1 GCCCGCTGG CCATCAAGTC CCTCACCGAA AGGTTTATO TTGGGGGCC TCTTACCAAT TCAAGGGGGG 2 1 1 GCCCGCTGG CCATCAAGTC CCTCACTGAA AGGTTTATO TTGGGGGCC TCTTACCAAT TCAAGGGGGG 3 61 71 GCCCGCTGG CCATCAAGTC CCTCACTGAA AGGTTTATO TTGGGGGCC TCTTACCAAT TCAAGGGGG 4 61 71 GCCCGCTGG CCATCAAGTC CCTCACTGAA AGGCTTTATG TTGGGGGCC TCTTACCAAT TCAAGGGGG 5 61 71 GCCCGCTGG CCATCAAGTC CCTCACTGAA AGGCTTTATG TTGGGGGCC TCTTACCAAT TCAAGGGGG 6 71 GCCCGCTGG CCATCAAGTC CTCACTGAA AGGCTTTATG TTGGGGGCC TCTTACCAAT TCAAGGGGC 7 GCCGGTGG CCATCAAGTC CTCACTGAA AGGCTTTATG TTGGGGGCC TCTTACCAAT TCAAGGGGC 8 71 GCCGGTGG CCATCAAGTC CTCACAGAA GGGCTTTATG TTGGGGGCC TCTTACCAAT TCAAGGGGC 9 71 GCCGGCGGG CCATAAGTC GTCACAGAA GGGCTTTATA TCGGGGGCC TTCTACCAAT TCAAAGGGC 10 71 GCCGGCGGG CCATAAGGTC GTCACAGAA GGGCTTTATA TCGGGGGCC CTCTACCAAT TCAAAGGGC 10 71 GCCGGCGGG CCATAAGGTC GTCACAGAA GGGCTTTATA TCGGGGGCC CTCTACCAAT TCAAAGGGC 10 71 GCCGGCGGG CCATAAGGTC GTCACAGAA GGGCTTTATA TCGGGGGCC CTCTACTAAT TCAAAGGGC 11 71 GCCGGCGGG CTATAAGGTC GTCACAGAA GGGTTTATA TCGGGGGCC CTGACTAAT TCAAAGGGC 12 71 GCCCGGAGG CTATAAGGTC GTCACAGAA GGGTTTATA TCGGGGGCC CTGACTAAT TCAAAGGGC 13 71 GCCCGAACTG CTAAACAC GTCACAGAA GGGTTTATA TCGGGGGCC CTGACTAAT TCAAAGGGC 14 71 GCCCGAACTG CTAAACAC GTCACAGAA GGGTTTATA TCGGGGGCC CTGACTAAT TCAAAGGGC 15 71 GCCCGAACTG CTAAACAC GTCACACAA AGGATTATA TCGGGGGGCC CTGACTAAT TCAAAGGGCC 16 71 GCCCGAACTG CTAAACAC GTCACACAA AGGATTATA AGGATGGCC CTGACTAAT TCAAAGGGCC 17 GCCCGAACTG CTAAACAC GTCACACAGA GGGCTTTACA TGGGGGGCC CTGACTAAT TCAAAGGGCC 18 71 GCCCGAACTG CTAAACAC GTCACACAA GGCTTAACA AGCAAGGGCC 18 71 GCCCGAACTG CTAAACAC GTCACACACA GGCTTAACA AGCAAGGGCC 19 71 GCCCGAACTG CTAAACAC GTCACACACA GGCTTAACA AGCAAGGGCC 10 71 GCCCGAACTG CTAACACAC GTCACACACA GGCTTAACA ACAAGGGCC 10 71 GCCCGAACTG CTAACACAC GTCACACACA GGCTTAACA ACAAGGGCC 10 71 GCCCGAACTG CTAACACAC GTCACACACA GGCTTAACA ACAAGGGCC 10 71 GCCCGAACTG CTAACACAC GTCAC	SEQUENCE		l I	
1 01 11 GCCGGCTGG CATCAAGTC CTCACCGAG AGGCTTTATG TTGGGGGCC TCTACCAAT 2 01 11 GCCGGCTGG CATCAAGTC CCTCACGAG AGGCTTTATG TTGGGGGCCC TCTACCAAT 3 01 11 GCCGGCTGG CCATCAAGTC CCTCACGAG AGGCTTTATG TTGGGGGCCC TCTACCAAT 1 1 GCCGGCTGG CCATCAAGTC CCTCACGAG AGGCTTTATG TTGGGGGCCC TCTACCAAT 1 1 GCCGGCTGG CCATCAAGTC CCTCACGAG AGGCTTTATG TTGGGGGCCC TCTACCAAT 1 1 GCCGGTGG CCATCAAGTC CCTCACGAG AGGCTTTATG TTGGGGGCCC TCTACCAAT 1 GCCGGTGGG CCATCAAGTC CCTCACGAG AGGCTTTATG TTGGGGGCCC TCTACCAAT 1 GCCAGACAG CCATCAAGTC CCTCACGAG GGCTTTATG TTGGGGGGCC TCTACCAAT 1 GCCAGACAG CCATCAAGTC GCTCACAGAG GGCTTTATA TCGGGGGCC TCTACCAAT 1 GCCAGACAG CCATCAAGTC GCTCACAGAG GGCTTTATA TCGGGGGCC CTGACCAAT 1 GCCAGACAG CCATCAAGTC GCTCACAGAG CGCTTTATA TCGGGGGCC CTGACCAAT 1 GCCAGACAG CCATCAAGTC GCTCACAGAG CGCTTTATA TCGGGGGCC CTGACCAAT 1 GCCAGACAG CATCAAGTC GCTCACAGAG CGCTTTATA TCGGGGGCC CTGACCAAT 1 GCCAGACAG CATCAACTC GCTCACAGAG CGCTTTATA TCGGGGGCC CTGACCAAT 1 GCCAGACAG CATCACACA GCTCACAGAG CGCTTTATA TCGGGGGCC CTGACCAAT 1 GCCAGACAG CATCACACA GCTCACAGAG CGCTTTATA TCGGGGGCC CTGACCAAT 1 GCCAGACAG CATCACACA GCTCACAGAG CGCTTTATA TCGGGGGCC CTGACACAC CTCACACAG CGCTGTAAA 1 GCCAGACAG CATCACACA GCTCACAGA GGCTTTATA TCGGGGGCC CTGACACAC CATCACACAC GCTCACACAG CGCTTAAA GCTCACACA GCCTGTAAA GCTCACACA GCCTGTAAA GCTCACACA GCCTGTAAA GCTCACACA GCCTGTAAA GCGGGGCC CATGACAAA GCTCACACA GCCTGAAA GCTCACACA GCCTGAAAA GCTCACACA GCCTGAAAA GCTCACACA GCCTGAAAA GCTCACACA GCCTGAAAA GCTCACACA GCCGCGGGCC CATGACAAA GCTCACACA GCCTGAAAA GCTCACACA GCCTGAAAA GCTCACACA GCCTGAAAA GCTCACACA GCCTGAAAA GCTCACACA GCCCTGAAAA GCTCACACA GCCTGAAAA GCTCACACA GCCGCGCC CATGACACA GCCCACACACA GCCCACACACA GCCCACACACA	ID NUMBER	GENOTYPE		
01 71 GCCGGGTGG CCATCAGGT CCTCACGGA AGGTTTATG TTGGGGGCC TCTTACCAAT 2 1 GCCGGGTGG CCATCAGGTG CCTCACGGA AGGTTTATG TTGGGGGCCC TCTTACCAAT 3 1 GCCCGGTGG CCATCAGGT CCTCACTGA AGGTTTATG TTGGGGGCCC TCTTACCAAT 4 1 GCCCGGTGG CCATCAGGT CCTCACTGA AGGTTTATG TTGGGGGCC TCTTACCAAT 5 G1 71 GCCCGGTGG CCATCAGGT CCTCACTGA AGGTTTATG TTGGGGGCC TCTTACCAAT 6 T 1 GCCCGGTGG CCATCAGGT CCTCACTGA AGGTTTATG TTGGGGGCC TCTTACCAAT 7 GCCCGGTGG CCATCAGGT CCTCACTGA AGGTTTATG TTGGGGGCC TCTTACCAAT 6 T 1 GCCCGGTGG CCATCAGGT CCTCACTGA GGCTTTATG TTGGGGGCC TCTTACCAAT 7 GCCAGCAGG CCATCAGGT GCTCACAGG GGCTTTATG TTGGGGGCC TCTTACCAAT 6 T 1 GCCAGCAGG CCATCAGGT GCTCACAGG GGCTTTATA TCGGGGGCC CCTGACTAAT 7 GCCAGCAGG CCATCAGGT GCTCACAGG GGCTTTATA TCGGGGGCC CCTGACTAAT 1 GCCAGCAGG CCATCAGGT GCTCACAGG GGCTTTATA TCGGGGGCC CCTGACTAAT 1 GCCAGCAGG CCATCAGGT GCTCACAGG GGCTTTATA TCGGGGGCC CCTGACTAAT 1 GCCAGCACGG CTATCAGGT GCTCACCAGG GGCTTTATA TCGGGGGCC CCTGACTAAT 1 GCCAGCACG CTATCAGGT GCTCACCAGG GGCTTTATA TCGGGGGCC CCTGACTAAT 1 GCCAGCACG CTATCAGGT GCTCACCAGG GGCTTTATA TCGGGGGCC CCTGACTAAT 1 GCCAGCACG CTATCAGGT GCTCACCAGG GGCTTTACA TCGGGGGCC CCTGACTAAT 1 GCCCGAACAG GCTCACACGG GCTCACACG GGCTTTACA TCGGGGGCC CCTGACTAAT 1 GCCCGAACAG GCTCACACT GCTCACAGG GGCTTTACA TCGGGGGCC CTGACTAAT 1 GCCCGAACAG GCTCACACT GCTCACAGG GGCTTTACA TCGGGGGCC CTGACACAG GCTCACACT GCTGACTAAG AGGTTTACA TGGGGGCC CTGACACAG GCTCACACT GCTGACTAGA AGGTTTACA TGGGGGGCC CTGACACAG GCTCACACT GCTGACTACA GGCCGTTACACT GTGGGGGCC CTGACACAG GCTCACACT GCTGACTACA GGCCGTTACACAC GCTCACACT GTGGGGGCC CTGACACAC GCTCACACT GCTGACGGGCC CTGACACAG GCTCACACT GCTGACGGCC CTGACTACAC GCTCACACT GCGGGGGCC CTGACACAC GCTCACACAC	# # # # # # #	85 11 10 10 10 11 11 11	11 11 11	
2 GI 71 GCCGCGTGG CCATCAAGTC CCTCACTGAG AGGCTTTATG TCGGGGGCC TCTRACCAAT 3 4 GI 71 GCCGGGTGG CCATCAAGTC CCTCACTGAG AGGCTTTATG TTGGGGGCCC TCTRACCAAT 5 5 GI 71 GCCGGGTGG CCATCAAGTC CCTCACTGAG AGGCTTTATG TTGGGGGCCC TCTRACCAAT 5 6 GI 71 GCCGGTGG CCATCAAGTC CCTCACTGAG AGGCTTTATG TTGGGGGCCC TCTRACCAAT 7 7 GCCGGTGG CCATCAAGTC CCTCACTGAG AGGCTTTATG TTGGGGGCCC TCTRACCAAT 7 8 71 GCCGGTGG CCATCAAGTC CCTCACTGAG AGGCTTTATG TTGGGGGCCC TCTRACCAAT 7 9 71 GCCGGCAACAG CCATCAAGTC CCTCACTGAG AGGCTTTATG TTGGGGGCCC TCTRACCAAT 7 10 GCCGGCAACAGG CCATCAAGTC GCTCACAGAG CGGCTTTATA TCGGGGGCCC TTGACTAAT 7 11 GCCAGACAGG CCATCAAGTC GCTCACAGAG CGGCTTTATA TCGGGGGCCC CTGACTAAT 7 12 GCCAGCAAGG CCATCAAGTC GCTCACAGAG CGGCTTTATA TCGGGGGCCC CTGACTAAT 7 13 GCCAGCAAGG CATCAAGTC GCTCACAGAG CGCTTTATA TCGGGGGCCC CCTGACTAAT 7 14 GCCAGCAAGG CTATAAGGTC GCTCACAGAG CGCTTTATA TCGGGGGCCC CCTGACTAAT 7 15 GCCAGCAAGG CTATAAGGTC GCTCACAGAG CGCTTTATA TCGGGGGCCC CCTGACTAAT 7 16 GCCAGCAAGG CTATAAGGTC GCTCACAGAG CGCTTTATA TCGGGGGCC CCTGACTAAT 7 17 GCCAGCAAGG CTATAAGGTC GCTCACAGAG CGCTTTAATA TCGGGGGCC CCTGACTAAT 7 18 GCCGAACAGG CTATAAGGTC GCTCACAGAG CGCTTTAATA TCGGGGGCC CCTGACTAAT 7 19 GCCGAACAGG CTATAAGGTC GCTCACAGAG CGCTTTAATA TCGGGGGCC CCTGACTAAT 7 10 GCCGAACAGG CTATAAGGTC GCTCACAGAG CGCTTTAATA TCGGGGGCC CCTGACTAAT 7 10 GCCGAACAGG CTATAAGGTC GCTCACAGAG CGCTTTAATA TCGGGGGCC CCTGACTAAT 7 10 GCCAACAAGG CTATAAGGTC GCTCACAGAG CGCTTTAATA TCGGGGGCC CCTGACTAAT 7 11 GCCAACAAGG CTATAAGGTC GCTCACAGAG CGCTTTAATA TCGGGGGCC CCTGACTAAT 7 11 GCCAACAAGG CTATAAGGTC GCTCACAGAG CGCTTTAATA TCGGGGGCC CCTGACTAAT 7 11 GCCAACAATG CCATAACACT GCTGACTAAG AGGCTTTAAT TCGGGGGCC CTGACTAAT 7 11 GCCCGAACAGG CTATAACAGTC GCTGACTAAG TGGGGGCC CTGACTAAT 7 12 GCCGAACAGG CTATAACACT GCTGACTAAG TGGGGGCC CATGACAAAC 7 13 GCCGAACAGG CATACACT GCTGACTAAG AGGCTGTAAC TGGGGGGCC CATGACAAAC 7 14 TTAAAA AAAA AAAA AAAAA AAAAA AAAAAAAAA	-	GI	7.1	CCGCGTGG CCATCAAGTC CCTCACCGAG AGGCTTTATG TTGGGGGCCC TCTTACCAAT
3 GI 71 GCCGGGGG CATCAAGTC CCTCACTGAG AGGCTTTATCG TTGGGGGCCC TCTTACCAAT 5 GI 71 GCCGGGTGG CCATCAAGTC CCTCACTGAG AGGCTTTATG TTGGGGGCCC CCTTACCAAT 6 GI 71 GCCGGGTGG CCATCAAGTC CCTCACTGAG AGGCTTTATG TTGGGGGCCC TCTTACCAAT 7 GCCGGTGG CCATCAAGTC CCTCACTGAG AGGCTTTATG TGGGGGCCC TCTTACCAAT 7 GCCGGTGG CCATCAAGTC CCTCACTGG AGGCTTTATG TGGGGGCCC TCTTACCAAT 7 GCCGGTGG CCATCAAGTC CCTCACTGG CGCTCTTATG TGGGGGCCC TCTTACCAAT 7 GCCAGACAG CCATAAGTC GCTCACAGG CGCTCTTATG TCGGGGGCCC CCTGACTAAT 7 GCCAGACAG CCATAAGTC GCTCACAGG CGCTCTTATA TCGGGGGCCC CCTGACTAAT 7 GCCAGACAG CTATAAGTC GCTCACAGG CGCTTTATA TCGGGGGCCC CCTGACTAAT 7 GCCAGACAG CTATAAGTC GCTCACAGG CGCTTTACA TCGGGGGCCC CCTGACTAAT 7 GCCCGAACTG CTATACACT ACTGACTAAG CGCTTTACA TCGGGGGCC CTTGACAAC 7 GCCCGAACTG CTATACACT ACTGACTGAG AGACTTATAC TGGGGGGCC CATGACAAC 7 GCCCGAACTG CTATACACT ACTGACTGAG AGACTTATAC TGGGGGGCC CATGACAAC 7 GCCCGAACTG CTATACACT ACTGACTGAG AGACTTATAC TGGGGGGCC CATGACAAC 7 GCCCGAACTG CATACACTG ACTCACTGAG AGACTTATAC TGGGGGGCC CATGACAAC 7 GCCCGAACTG CATACACTG ACTCACTGAG AGACTCTACC ACTGACAAC 7 GCCCGAACTG CATACACTG ACTCACTGAG AGACTCTACC ACTGACAAC CATCACTGA CATCACTACA CATCACTACA CATCACTACA CATCACTACA CATCACTACA CATCACTACA CATCACTACA CATCACTACACA CATCACTACACA CATCACTACACA CATCACTACACA CATCACTACACA CAT	~	GI	7.1	CCGCATGG CCATCAAGTC CCTCACTGAG AGGCTTTATG TCGGGGGCCC TCTTACCAAT
GI 71 GCCGGGGG CCATCAAGTC CCTCACTGAG AGGCTTTATG TGGGGGCC CCTTACCAAT	٣	ĠĬ	7.1	CCGCGTGG CCATCAAGTC CCTCACTGAG AGGCTTTACG TTGGGGGCCC TCTTACCAAT '
5 61 71 GCCGGGGG CCATCAAGTC CCTCACGGA AGGCTTTATG TCGGGGGCC TCTTACCAAT	ぜ	19	17	CCGCGIGG CCAICAAGIC CCICACIGAG AGGCITIAIG IIGGGGGCCC CCTIACCAAI
6 GI 71 GCCGTGTGG CCATCAGGT CCTCACTGG AGGCTTTATG TGGGGGCC TCTTACCAT 7 GCCAGCAGG CCATAGGT GCTCACAGG CGGCTTTATG TCGGGGGCC CTTACCATA 1 GCCAGCAGG CCATAGGT GCTCACAGG CGGCTTTATA TCGGGGGCC CTGACTAT 1 GCCAGCAGG CCATAGGT GCTCACAGG CGGCTTATA TCGGGGGCC CTGACTAT 1 GCCAGCAGG CTATAGGT GCTCACAGG GGCTTTATA TCGGGGGCC CTGACTAT 1 GCCAGCAGG CTATAGGT GTCACAGG GGCTTTATA TCGGGGGCC CTGACTAT 1 GCCAGCAGG CTATAGGT GCTCACAGG GGCTTTACA TCGGGGGCC CTGACTAT 1 GCCCGACAG CTATAGGT GTCACACT GTGGCGGGCC CTGACTAT 1 GCCCGACAG CTATACACT GTGGCTGTAG AGCTTTACA TGGGGGGCC CTGACTAAC 1 GCCCGACAG CTATACACT GTGGCTGTAG AGCTTTACA TGGGGGGCC CTGACTAAC 1 GCCCGACAG CTATACACT GTGGCTGAG GACTTACA TGGGGGGCC CTGACTAAC 1 GCCCGACAG CTATACACT GTGGCTGAG GACTTACA TGGGGGGCC CATGACAAC 1 GCCGAACAG CTATACACT GTGACTGAG GACTTACA TGGGGGGCC CATGACAAC 1 GCCGAACAG CTATACACT GTGACTGAG GACTTACA TGGGGGGCC CATGACAAC 1 GCCGAACAG CTATACACT GTGACTGAG GACTTACA GTGGGGGCC CATGACAAC 1 GCCGGAACT GTATACACT GTCACCCAA GCCTTACA TGTGACAAC 1 GCCGGAACT GTATACACT GTCACCCAA GCCTTACA TGTGAGGCC CATGACAAC 1 GCCGGAACA GCCTCACGAA GCCTTACA TGTGAGGCC CATGACAAC 1 GCCGGAACA GCCTCACGAA GCCTTACA TGTGAGGCC CATGACAAC 1 GCCGGAACA GCCTCACGAA GCCTTACA CGCTGAAC GCCTGAACA GCCTTACA CGCTGAACA CGCTACAACA CGCTACAACA CGCTACAACA CGCTACAACA CGCTACAACA CGCTACAACA	2	GI	7.1	CCGCGTGG CCATCAAGTC CCTCACCGAG AGGCTTTATG TCGGGGGCCC TCTTACCAAT '
	•	GI	7.1	CCGIGIGG CCAICAAGIC CCICACIGAG AGGCITIAIG IIGGGGGCCC ICTIACCAAI
GII 71 GCCAGACAG CCATAAGGT GCTCACAGAG GGCTTTATG TCGGGGGCC CTGACTAAT	11 11 11 11 11 11	11 11 11 11 11 11	11 11 11 11	
1 GCCAGACAAG CCATAAGGTC GCTCACAGAG CGGCTTTATA TCGGGGCCC CCTGACTAAT 1 GCCAGCAGG CCATAAGGTC GCTCACAGAG CGGCTTTATA TCGGGGGCC CCTGACTAAT 1 GCCAGCAGG CCATAAGGTC GCTCACCAGG CGGCTTTATA TCGGGGGCC CCTGACTAAT 1 GCCAGCAGG CTATAAGGTC GCTCACCAGG CGGCTTTATA TCGGGGGCC CCTGACTAAT 1 GCCAGCAGG CTATAAGGTC GCTCACCAGG CGCTTTAATA TCGGGGGCC CCTGACTAAT 1 GCCAGCAGG CTATAAGGTC GCTCACCAGG CGCTTTAATA TCGGGGGCC CCTGACTAAT 1 GCCAGACAGG CTATAAGGTC GCTCACCAGG CGCTCTAATA 1 GCCCGAACTG CTATAAGGTC GCTCACCAGG CGCTCTAATA 1 GCCCGAACTG CTATACACTC ACTGACTGAG AGGCTCTAAC GTGGGGGCC CATGACAAAC 1 GCCCGAACTG CATACACTC ACTGACTGAG AGGCTCTAAC GTGGGGGCC CATGACAAAC 1 GCCCGAACTG CATACACTC ACTGACTGAG AGGCTCTAAC GTGGGGGCC CATGACAAAC 1 GCCCGAACTG CATACACTC ACTGACTGAG AGGCTCTAAC GTGGGGGCC CATGACAAAC 1 GCCGGGGGC CATCACCACC GCTCACCCAA GGCTGTACC CATGACAAAC 1 GCCGGGGGC CATCACCACC GCTCACCCCAA GGCTGTACC CATGACAAAC 1 GCCGGGGGC CATCACCCCAA GGCTGTACC GTGGGGGCC CATGACAAAC 1 GCCGGGGG CAATACGGTC ACTCACCCAA GGCTGTACC GTGGGGGCC CATGACAAAC 1 GCCGGGGG CAATACGGTC ACTCACCCAA GGCTGTACT GTGGGGGCC CATGACAAAC 1 GCCGGGGGG CAATACGGTC ACTCACCCCAA GGCTGTACT GTGGGGGCC CATGATATAAC 1 GCCGGGGGG CAATACGGTC ACTCACCCCAA GGCTGTACT GTGGGGGCC CATGATATAAC 1 GCCGGGGGG CAATACGGTC CCTCACGGGA GGCTCTACT GTGGGGGCC CATGATATAAC 1 GCCGGGGGG CAATACGGTC CCTCACGGGG CGCTGTACCT ACTGATATAAC GGCTGTACCAAC GGCTGTACCAAC GGCTGTACCAAC GGCTGTACCAAC GGCTGTACAAC	7	119	71	CAGACAGG CCATAAGGTC GCTCACAGAG CGGCTCTATG TCGGGGGTCC TATGACTAAC
1 GCTGGCAGG CCATAAGGT GCTCACAGA GGCTTTATA TCGGGGCCC CCTGACCAAT 1 GCCAGCAGG CCTTAAGGT GCTCACAGA GGCTTTATA TCGGGGGCC CCTGACTAAT 1 GCCAGACAGG CTATAAGGT GCTCACAGA GGCTTTACA TCGGGGGCC CCTGACTAAT 1 GCCAGACAGG CTATAAGGT GCTCACAGA GGCTTTACA TCGGGGGTC CCTGACTAAT 1 GCCCGACAGG CTATAAGGT GCTCACAGA GGCTTTACA TCGGGGGTC CCTGACTAAT 1 GCCCGACAGG CTATACACT GCTGACTAGG GGCTCTACG TGGGGGGC CTTGACTAAC 1 GCCCGACTG CTATACACT GCTGACTAGG GGCTTTACA TAGGGGGC CTTGACAACC 1 GCCCGAACTG CTATACACT GCTGACTAGG GGCC CTTGACAAACC 1 GCCCGAACTG CTATACACTG GCTGACTAAC GTGGGGGC CATGACAAACC 1 GCCCGAACTG CTATACACTG GCTGACTAACCGT GTGGGGGC CATGACAAACC 1 GCCCGAACTG CTATACACTG GCTCACTGA GGCTGTACT GTGGAGGCC CATGACAAACC 1 GCCCGAACTG CATACAGTG AGCTGTACT GTGGAGGCC CATGACAAACC 1 GCCCGTGTG CATACGGT ACTCACCCAA GGCTGTACT GTGGAGGCC CATGACAAACC 1 GCCGGTGTG CATACGGT ACTCACCCAA GGCTGTACT GTGGAGGCC CATGACAAACC 1 GCCGGGGGG CAATACGGT ACTCACCCAA GGCTGTACT GTGGAGGCC CATGACAACC 1 GCCGGGAAAG GATCCCCCAA GGCTGTACT GGGGGGCC CATGACAACC 1 GCCGGGAAAG TGATCCCCCA CCTCACGGAA GGCTTACT GCGGGGCC TATGTTCAACC 1 GCCAGGAAAG TGATCCCTC CCTCACGGAA GGCTTACT GCGGGGCC TATGTTCAACC 1 GCCAGGAAAG TGATCCCTC CCTCACGAA CGCCTGTACT GCGGGGCC TATGTTCAACC 1 GCCAGGAAAG TGATCCCTC CCTCACGAACC CCTCACACACC CACCACACACC CCTCACACACC CCTCACACACC CCTCACACACC CCTCACACACC CACCACACACC CCTCACACACC CCTCACACACC CCTCACACACC CACCACACACC CCTCACACACC CACCACACACC CACCACACACC CACCACACACC CACCAC	8		71	CAGACAAG CCATAAGGTC GCTCACAGAG CGGCTTTACA TCGGGGGCCC CCTGACTAAT
10 71 GCCAGGCAGG CCATAAGGT GCTCACCGAG GGACTTATA TCGGGGGCC CCTGACTAAT 71 GCCAGACAGG CTATAAGGT GCTCACAGAG GGGTGTACA TCGGGGGTC CCTGACTAAT 71 GCCAGACAGG CTATAAGGT GGTCACTAGA GGGTTTACA TCGGGGGTC CCTGACTAAT 71 GCTCACATT GCTGACTGAG AGGTTTACG TGGGGGGC CATGACAAC 71 GCTCAACATT GCTGACTGAG AGGTTTACG TAGGGGGC CATGACAAC 71 GCTCGAACTG CTATACACT ACTGACTGAG AGGTTTACG TAGGGGGC CATGACAAC 71 GCTCGAACTG CTATACACT ACTGACTGAG AGGTTTACG TAGGGGGC CATGACAACC 71 GCTCGAACTG CTATACACT ACTGACTGAG AGGTTTACG TAGGGGGC CATGACAACC 71 GCTCGAACTG CTATACACTC ACTGACTGAG AGGTTTACG TAGGGGGC CATGACAACC 71 GCTCGAACTG CTATACACTC ACTGACTGAG AGGTTTACG TAGGGGGCC CATGACAACC 71 GCTCGAACTG CTATACACTC ACTGACTGAG AGACTGTACG TAGGGGGCC CATGACAACC 71 GCTCGAACTG CTATACACTC ACTGACTGAG AGACTCTACG TAGGGGGCC CATGACAACC 71 GCTGGAACTG CTATACACTC ACTGACCCAA CGCTGTACT GTGGAGGCC CATGACAACC 71 GCTGGGGGC CATACGGTC ACTCACCCAA CGCTGTACT GTGGAGGCC CATGATAACC 71 GCCGGGGGG CAATACGGTC ACTCACCCAA CGCTGTACT GTGGAGGCC CATGATAACC 71 GCCGGGGGG CAATACGGTC ACTCACCCAA CGCTGTACT GTGGAGGCC CATGATAACC 71 GCCAGGAAG TGATCCCTC CCTCACGGAG CGCTGTACT GCGGGGGCC TATGTTCAACC 71 GCCAGGAAAG TGATCCCTC CCTCACGGAG CGCTTTACT GCGGGGCC TATGTTCAACC 71 GCCAGGAAAG TGATCCCTC CCTCACGGAA CGCTTTACT GCGGGGCCC TATGTTCAAT 71 GCCAGGAAAG TGATCCCTC CCTCACGGAA CGCTTTACT GCGGGGCCC TATGTTCAAT 71 GCCAGGAAAG TGATCCCTC CCTCACGGAA CGCCTTAACT GCGGGGCC TATGTTCAAT 71 GCCAGGAAAG TGATCCCTC CCTCACGGAA CGCCTTAACT GCGGGGCC TATGTTCAAT 71 GCCAGGAAAG TGATCCCTC CCTCACGGAA CGCCTTAACT CCTCACGGAA CGCCTTAACC CATGACAACC CTCACGGAA CGCCTTAACC CATGACAACC CTCACGGAA CGCCTTAACCAC CATGACAACC CTCACGGAA CGCCTTAACCAC CATGACAACC C	6		7.1	TAGACAGG CCATAAGGTC GCTCACAGAG CGGCTTTATA TCGGGGGCCC CCTGACCAAT
11 71 GCCAGACAGG CTATAAGGTC GCTCACAGAG CGGCTGTACA TCGGGGGTCC CCTGACTAAT 1 GCCAGACAGG CTATAAGGTC GCTCACAGAG CGGCTTTACA TCGGGGGTCC CCTGACTAAT 1 GCCACATTG CCATACACTG GCTGACTAACG TCGGGGGTCC CATGACTAACG CTATACACTG CTGACTGAG AGGCTCTACG TGGGGGGCC CATGACAACG CTGACACACG TAGACAGGGCC CATGACAACG TAGACAGGGGCC CATGACAACG TAGACAGGGGCC CATGACAACG TAGACAGGGGCC CATGACAACG TAGACAGGGGCC CATGACAACG TAGACAGGGGCC CATGACAACG TAGACAGGGGCC CATGACAAACG TAGACACGTG CATGACTACACGTG AGGCTGTACG TAGGGGGGCC CATGACAACG TAGACAGGGCC CATGACAAACGTG CATACACTG CATGACTACG TAGGGGGGCC CATGACAACGGTG CATACACTG CATGACTGTAG AGGCTGTACG TAGGGGGCC CATGACAAACGTG CATACACTG CATGACTACGGTG CATGACAAACGGTG CATACACTG CATCACCCAA GCCTGTACTACG TAGGGGGCC CATGACAAACGGTG CATACACTG CATCACCCCAA GCCTGTACT GTGGAGGCC CATGACAAACGGTG CATACACCTG CATCACCCCAA GCCTGTACT GTGGAGGCC CATGATAAACGGTC CATCACCCCAA GCCTGTACT GTGGAGGCC CATGATAAACGGTC CATCACCCCAA GCCTGTACT GTGGAGGCCC CATGATAAACGGTC CATCACCCCAA GCCTGTACT GTGGAGGCCC CATGATAAACGGTC CATCACCCCAA GCCTGTACT GTGGAGGCCC CATGATAAACGGTC CATCACCCCAA GCCTGTACT GTGGAGGCCC TATGTTCAACGGTC CATCACGGAA GGCTTTACT GCGGGGCCC TATGTTCAAT GCGCGGGCC TATGTTCAAT GCGGGGCCC TATGTTCAAT GCGGGGCCC TATGTTCAATCAACGGTAACCTCCTCCCCCCACGGGAA GGGCTTTACTTCATCAACCTCCCCCACGGAA CGGCTTTACTTCATCACCCACGGAA CGGCTTTACTTCAATCAACCTCACGGAA CGGCTTTACTTCAATCAACCTCACGAAAC CCCCCACGCGGAA CGGCTTTACTTCAACCTCACGAAAC CCCCACACGAAACCTCAACACCAACCAACCAACCAACCAA	10		11	CAGGCAGG,CCATAAGGTC GCTCACCGAG CGACTTTATA TCGGGGGCCC CCTGACTAAT
12 71 GCCAGACAGG CTATAAGGTC GCTCACAGAG CGGCTTTACA TCGGGGGTCC CCTGACTAAT 1 13 GIIF 71 GCTCACATTG CCATACACTC GCTGACTGAG AGGCTCTACG TGGGGGGCC CATGTTCAAC 14 71 GCTCGAACTG CTATACACTC ACTGACTGAG AGACTATACG TAGGGGGCC CATGACAAAC 15 71 GCTCGAACTG CTATACACTC ACTGACTGAG AGACTGTACG TAGGGGGCC CATGACAAAC 17 GCTCGAACTG CTATACACTC ACTGACTGAG AGACTGTACG TAGGGGGCC CATGACAAAC 18 GV 71 GCTCGAACTG CTATACACTC ACTGACTGAG AGACTGTACG TAGGGGGCC CATGACAAAC 19 GCTCGAACTG CTATACACTC ACTGACTGAG AGACTGTACG TAGGGGGCC CATGACAAAC 10 GCTCGAACTG CTATACACTC ACTGACTGAG AGACTGTACG TAGGGGGCC CATGACAAAC 11 GCTCGAACTG CTATACACTC ACTCACCCCAA CGCCTGTACT GTGGAGGCC CATGACAAAC 11 GCTCGAACTG CAATACGGTC ACTCACCCCAA CGCCTGTACT GTGGAGGCC CATGTATAAC 11 GCCGGGGAAG TAATACGGTC ACTCACCGCAA CGCCTGTACT GTGGAGGCC CATGTATAAC 12 GCAGGGAAAG TGATCTCCTC CCTCACGGAG CGCCTTTACT GCGGGGCCC TATGTTCAAT 12 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGCCTTTACT 12 GCCAGGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACT 13 GCCAGGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACT 14 GCCAGGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACT 15 GCCAGGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACT 15 GCCAGGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACT 16 GCCAGCAAAC 17 GCCAGGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACT 17 GCCAGGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACTTACT 18 GCCAGGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACT 18 GCCAGCTTTACT 18 GCCAGCAAAC 19 GCCAGCACTACACCT 19 GCCAGCAAAC 10 GCCCAAACACT 10 GCCCAAACACT 10 GCCCAAACACT 10 GCCCAAACACT 10 GCCCAAACACT 10 GCCCCAAACACT 11 GCCCAAACACT 11 GCCCCAAACACT 11 GCCCAAA	11		7.1	CAGACAGG CTATAAGGTC GCTCACAGAG CGGCTGTACA TCGGGGGTCC CCTGACTAAT
13 GIIF 71 GCTCACATTG CCATACACTC GCTGACTGTACG TGGGGGGCC CATGTCAACC 14 T1 GCTCCAACTG CTATACACTC ACTGACTGTAG AGGCTCTACG TGGGGGGCC CATGACAAAC 15 T1 GCCCGAACTG CTATACACTC ACTGACTGTAG AGACTGTACG TAGGGGGGC CATGACAAAC 17 GCTCGAACTG CTATACACTC ACTGACTGTAG AGACTGTACG TAGGGGGGC CATGACAAAC 17 GCTCGAACTG CTATACACTC ACTGACTGTAG AGACTGTACG TAGGGGGGC CATGACAAAC 18 GCTCGAACTG CTATACACTC ACTGACTGTAG AGACTGTACG TAGGGGGGC CATGACAAAC 18 GCTCGAACTG CTATACACTC ACTGACTGTAG AGACTCTACG TAGGGGGGC CATGACAAAC 18 GCTAGAACTG CTATACACTC ACTGACTGTAG AGACTCTACG TAGGGGGCC CATGACAAAC 18 GCTAGAAACTG CATACGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCC CATGATAAC 19 GCAGGGAGG CAATACGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCC CATGTATAAC 19 GCAGGGAAG TAATCGGTC ACTCACCGAA CGCCTGTACT GTGGAGGCCC TATGTTCAAC 20 GIV 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAT 21 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAT 22 T1 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAT 22 T1 GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGGCTTTACT GCGGGGGCCC TATGTTCAAT 22 T1 GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGGCTTTACT GCGGGGGCCC TATGTTCAAT CCTCACGGAA CGGCTTTACT GCGGGGGCCC TATGTTCAAT CCTCACGGAA CGGCTTTACT GCGGGGCCC TATGTTCAAT CCTCACGGAA CGGCTTTACT GCGGGGGCCC TATGTTCAAT CCTCACGGAA CGGCTTTACT GCGGGGGCCC TATGTTCAAT CCTCACGGAA CGCCTTTACT GCGGGGGCCC TATGTTCAAT CCTCACGGAA CGCCTTTACTTCAAT GCGGGGGCCC TATGTTCAAT CCTCACGGAA CGCCTTTACT GCGGGGCCC TATGTTCAAT CCTCACGGAA CGCCTTTACTTCAAT CCTCACGGAA CGCCTTTACTTCAAT CCTCACGGAA CGCCTTTACTTCAAT CCTCACCGAA CGCCTTTACTTCAAT CCTCACCGAA CGCCTTTACTTCAAT CCTCACCGAA CGCCTTTACTTCAAT CCTCACCGAA CGCCTTTACTTCAAT CCTCACCGAA CGCCTTTACTTCAAT CCTCACCACAA CCTCACACAA CCT	12		7.1	CAGACAGO CTATAAGGIC GCICACAGAG CGGCITIACA ICGGGGGICC CCIGACIAAI
13 GIIF 71 GCTCACATTG CCATACACTC GCTGACTGGG AGGCTCTACG TGGGAGGGCC CATGTTCAAC 14 71 GCTCGAACTG CTATACACTC ACTGACTGAG AGACTATACG TAGGGGGCC CATGACAAC 15 71 GCCCGAACTG CTATACACTC ACTGACTGAG AGACTGTACG TAGGGGGCC CATGACAAC 16 71 GCTCGAACTG CCATACACTC ACTGACTGAG AGACTGTACG TAGGGGGCC CATGACAAC 17 GCTCGAACTG CCATACACTC ACTGACTGG TAGGGGGCC CATGACAAC 18 GV 71 GCTAGAACTG CTATCCACTC GCTCACTGG TAGGGGGCC CATGACAAC 19 T1 GCCAGGAACTG CATACGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCC CATGATAAC 19 T1 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGCTGTACT GTGGAGGCC TATGTTCAAC 20 GIV 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGCTTTACT GCGGGGGCCC TATGTTCAAC 21 T1 GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACT GCGGGGGCCC TATGTTCAAC 22 T1 GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACT GCGGGGGCCC TATGTTCAAC 22 T1 GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACT GCGGGGGCCC TATGTTCAAC	H H H H H H	61 61 61 12 63 13 13 14 14 18	H H U	
14 71 GCTCGAACTG CTATACACTC ACTGACTACG TAGGGGGCC CATGACAAAC 15 71 GCCCGAACTG CTATACACTC ACTGACTGACG TAGGGGGCC CATGACAAAC 16 71 GCTCGAACTG CTATACACTC ACTGACTGAG AGACTGTACG TAGGGGGCC CATGACAAAC 17 GCTCGAACTG CTATCCACTC ACTCACTGAG AGACTGTACG TAGGGGGCC CATGACAAAC 18 GV 71 GCTAGAACTG CTATCCACTC ACTCACCCAA CGCTGTACT GTGGAGGCC CATGACAAC 19 71 GCGCGTGTGG CAATACGGTC ACTCACCCAA CGCTGTACT GTGGAGGCC CATGTATAAC 19 71 GCACGCGGG CAATACGGTC ACTCACCCAA CGCTGTACT GTGGAGGCC CATGTATAAC 20 GIV 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAC 21 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAC 22 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACT GCGGGGGCCC TATGTTCAAC 22 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGGCTTTACT GCGGGGGCCC TATGTTCAAC	13	II	7.1	TCACATTG CCATACACTC GCTGACTGAG AGGCTCTACG TGGGAGGGCC CATGTTCAAC
15 71 GCCCGAACTG CTATACACTC ACTGACTGAC AGACTGTACG TAGGGGGCC CATGACAAAC 16 71 GCTCGAACTG CCATACACTC ACTGACTGAG AGGCTGTACG TAGGGGGCC CATGACAAAC 17 GCTAGAACTG CTATCCACTC GCTCACTGAG AGACTCTACG TAGGGGGCC CATGACAAAC 18 GV 71 GCGCGTGTGG CAATACGGTC ACTCACCCAA CGCTGTACT GTGGAGGCC CATGTATAAC 19 71 GCACGCGCG CAATACGGTC ACTCACCCAA CGCTGTACT GTGGAGGCC CATGTATAAC 20 GIV 71 GCACGGAAAG TGATCTCCTC CCTCACGGAG CGCTTTACT GCGGGGGCC TATGTTCAAC 21 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGCTTTACT GCGGGGGCCC TATGTTCAAC 22 TT GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACT GCGGGGGCCC TATGTTCAAT 22 TT GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACT GCGGGGGCCC TATGTTCAAC 22 TT GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGGCTTTACT GCGGGGGCCC TATGTTCAAT	14		71	TCGAACTG CTATACACTC ACTGACTGAG AGACTATACG TAGGGGGCC CATGACAAAC
16 71 GCTCGAACTG CCATACACTC ACTGACTGAG AGGCTGTACG TAGGGGGCC CATGACAAAC 17 71 GCTAGAACTG CTATCCACTC GCTCACTGAG AGACTCTACG TAGGAGGCC CATGACAAAC 18 GV 71 GCGCGTGTGG CAATACGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCC CATGTATAAC 19 71 GCACGCGGG CAATACGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCC CATGTATAAC 20 GIV 71 GCACGGAAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAC 21 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGCCTTACT GCGGGGGCCC TATGTTCAAT 22 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGCCTTACT GCGGGGGCCC TATGTTCAAT 22 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACT GCGGGGCCC TATGTTCAAT 22 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGCCTTTACT GCGGGGCCC TATGTTCAAT	15		11	CCGAACTO CTATACACTC ACTGACTGAG AGACTGTACG TAGGGGGGC CATGACAAAC
17 71 GCTAGAACTG CTATCCACTC GCTCACTGAG AGACTCTACG TAGGAGGCC CATGACAAAC 18 GV 71 GCGCGTGTGG CAATACGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCC CATGTATAAC 19 71 GCACGCGGG CAATACGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCC CATGTATAAC 20 GIV 71 GCCAGGAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAC 21 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAT 22 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAT 22 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGGCTTTACT GCGGGGGCCC TATGTTCAAT	16		7.1	TCGAACTG CCATACACTC ACTGACTGAG AGGCTGTACG TAGGGGGGCC CATGACAAAC
18 GV 71 GCGCGTGTGG CAATACGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCCC CATGTATAAC 19 71 GCACGCGCG CAATACGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCCC CATGTATAAC 19 71 GCACGCGGG CAATACGGTC ACTCACCGAA CGCCTGTACT GTGGAGGCCC CATGTATAAC 20 GIV 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAC 21 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAT 22 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGGCTTTACT GCGGGGGCCC TATGTTCAAT	17		11	TAGAACTG CTATCCACTC GCTCACTGAG AGACTCTACG TAGGAGGGCC CATGACAAAC
19 71 GCACGCGGG CAATACGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCCC CATGTATAAC 20 GIV 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAC 21 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAT 22 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGGCTTTACT GCGGGGCCC TATGTTCAAT	11	"	71	CCGCGTGTGG CATACGGTC ACTCACCAA CGCCTGTACT GTGGAGGCC CATGTATAC
20 GIV 71 GCCAGGAAG TGATCTCCTC CCTCACGGAG CGGCTTACT GCGGGGCCC TATGTTCAAC 21 71 GCCAGGAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGCCC TATGTTCAAC 22 71 GCCAGGAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGCCC TATGTTCAAT 22 71 GCCAGGAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGCCC TATGTTCAAT	19		7.1	ACGCGCGG CAATACGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCCC CATGTATAAC
71 GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGCCTTTACT GCGGGGGCCC TATGTTCAAT 71 GCCAGGAAAG TGATCTCCTC CCTCACGGAA CGGCTTTACT GCGGGGGCCC TATGTTCAAC			23	COLINGANA TONTOCATION OF COLORAGES CONTRACTOR COLORAGES AND TANGANA COLORAGES AND TANGANA COLORAGES COLORA
71 GCCAGGAAG TGATCTCCTC CCTCACGGAA CGGCTTTACT GCGGGGCCCC TATGTTCAAC		÷		CASCARAS TRATCHECTE CETTEREGRAS CRACETARET REGRANGEE TATEMERATE
)			: -	CALGOARA TORTOTOTOTO CONTINUE CONTINUE COCCURATION COCCURATION CONTINUES
	77		1	cheenand tenicity cettingens coortinact ecoegoscic inicitions

Fig. 2c

NSS REGION - (3/5)

11 11 11 11 11 11 11				
и н 4 го о	8 6 8 8 8 9 9 9 9 9	141 141 141 141 141	AGAACTGCGG CTATCGCAGG TGCCGCGCGA GCGGCGTACT GACAACTAGCCCCCCCCCC	" " " " " " " " " " " " " " " " " " "
7 8 8 9 10 11 12 13 14 15	611	141 141 141 141 141 141 141 141	AGAACTGCGG CTATCGCCGG TGCCGCGCGA GCGGCGTGCT AGAACTGCGG CTATCGCCGG TGCCGCGCCA GCGGTGTGCT AGAACTGCGG TTATCGCCGG TGCCGCGCCA GCGGCGTACT AGAACTGCGG TTATCGCCGG TGCCGCGCGA GCGGCGTGCT AGAACTGCGG CTATCGCCGG TGCCGCGCAA GCGGCGTGCT AGAACTGCGG CTATCGCCGG TGCCGCGCAA GCGGCGTGCT AGAACTGCGG GTACAGGCGT TGCCGCGCCAA GCGGCGTGCT AATCCTGCGG GTACAGGCGT TGCCGCGCGA GCGGAGTGCT AATCCTGCGG GTACAGGCGT TGCCGCGCGA GCGGAGTGCT AATCCTGCGG TTACAGGCGT TGCCGCGCGA GCGGAGTGCT AATCCTGCGG TTACAGGCGT TGCCGCGCGA GCGGAGTGCT AATCCTGCGG TTACAGGCGT TGCCGCGCCA GCGGGGTTTT	CCCTCACATG CCCTCACATG CCCTCACATG CCCTCACATG CCCTCACATG CCCTCACATG CCTCACATG CCTCACATG CCTCACATG CATCACGTG CACTCACGTG CACTCACGTG CACTCACGTG CACTCACGTG
19 19 20 21 22	GV (1)	141 141 141 15 16 16 141 141	i	ACA CCATGACGTG ACA CCATGACGTG ACA CCATGACGTG ACA CCATGACGTG ACA CAATCACTTG ACA CAATCACTTG ACA CAATCACTTG

y. 2d

NSS REGION - (4/5)

TO NIMBER							
	GENOTYPE						
		211	CTACATCAAG GCCCGGGCAG CCTGTCGAGC CGCAGGGCTC CAGGACTGCA CCATGCTCGT GTGTGGCGAC	CGCAGGGCTC CAGGACTGCA CCATGCTCGT	AGGACTGCA C	CATGCTCGT (GTGTGGCGAC
7		211	CTACATCAAG GCCCGGGCAG CCTGTCGAGC C	CGCAGGGCTC CA	CAGGACTGCA C	CCATGCTTGT (GTGTGGCGAC
3		211	CTACATCAAG GCCCGGGCAG CCTGTCGAGC C	CGCAGGGCTC C(CGGGACTGCA C	CCATGCTCGT (GTGTGGTGAC
' ঝ'		211	CTACATTAAG GCCCGGGCAG CCTGTCGAGC Ö	CGCAGGGCTC	CAGGACTGCA C	CCATGCTCGT (GTGTGGCGAC
2		211	TTACATCAAG GCCCAAGCAG CCTGTCGAGC C	CGCAGGGCTC C	CGGGACTGCA C	CCATGCTCGT	GTGGCGAC
9		211		CCIGICGAGC CGCAGGGCIC CAGGÀCIGCA CCAIGCICGI	AGGÁCTGCA C	CATGCTCGT (GTGTGGCGAC
	ununununun GII	======= 211	CTACCTGAAG GCCACAGGGG CCTGTCGAGC TGCCAAGCTC CAGGACTGCA CGATGCTCGT GAACGGAGAC	TGCCAAGCTC C	AGGACTGCA C	GATGCTCGT (SAACGGAGAC
&		211	AAG	TGCGAAGCIC C	CAGGACTGCA C	CGATGCTCGT	GTGCGGAGAC
6		211	GCCTCTGCAG CCTGTCGAGC		CAGGACTGCA C		GTGTGGGGAC
10		211	TIACTIGAAG GCCICIGCAG CCIGICGAGC I	TGCAAAGCTC C	CAGGACTGCA C	CGAIGCICGI	GAACGGGGAC
11		211	GCCTCTGCGG CCTGTCGAGC	TGCGAAGCIC C	CAGGACTGCA C	CGATGCTCGT	GTGCGGTGAC
12		211	TTACCTGAAG GCCAGTGCGG CCTGTCGAGC 1	TGCGAAGCTC CAGGACTGCA	AGGACTGCA C	CAATGCTCGT	GIGCGGIGAC
11 11 11 13 14 15 16 16	11 11 11 11 11 11 11	11 11 11		0	0 0 11 11 11 11 11 11 11 11 11 11 11 11	11 11 11 11 11 11 11 11 11 11 11 11 11	
13	GIII	211	CTATGTAAAA GCCCTAGCGG CTTGCAAGGC TGCAGGGATA GTTGCACCCT CAATGCTGGT ATGCGGCGAC	TGCAGGGATA G	TIGCACCCI C	AATGCTGGT	ATGCGGCGAC
14		211	CTACGTAAAA GCCAGGGCGG CGTGTAACGC C	CGCGGGGATT G	GTIGCICCCA C	CCATGCTGGT	GIGCGGIGAC
15		211		CCCCCCCATT G	GTTGCTCCCA C	CCATGITGGI	GTGCGCGAC
16		211	CTACGTGAAA GCTAAAGCGG CATGTAACGC C	CCCCCCCATT G	GTTGCCCCCA C	CCATGTTGGT	GIGCECCEAC
11		211	CIACAIC	TGCAGGGATC G	GIGGACCCIA TCAIGCIGGI	rcatectest	
		211	CTACATTAAG GCITTAGCCT CCTGTAGAGC CGCAAAGCTC CAGGACTGCA CGCTCCTGGT GTGTGGTGAT	CGCAAAGCIC C	AGGACTGCA (derecteer	GTGTGGTGAT
19		211	CTACATCAAG GCTTCAGCCG CCTGTAGAGC TGCAAAGCTC CAGGACTGCA CGCTCCTGGT GTGTGGTGTG	TGCAAAGCTC (AGGACTGCA (cerecreer	Grerere
20	GIV	211	TTACATCAAG GCTAGAGGG CTTCGAAGGC CGCAGGCCTC CGGAACCCGG ACTTTCTTGT CTGCGAAAT	CGCAGGCCTC	GGAACCCGG	ACTITICITGE	CIGCGGAGAT
2.1		211	TTACATCAAG GCTAGAGCGG CTGCGAAGGC CGCAGGGCTC		CGGACCCCGG ACTITCTCGT	ACTITICICGI	CTGCGGAGAT
22		211	TTACATCAAA GCTAGAGCGG CTGCCGAAGC CGCAGGCCTC		CGGAACCCGG	TAPATOTOTOTOTOTOTOTOTOTOTOTOTOTOTOTOTOTO	エタじてじじっじょう

Fig. 2e

NSS REGION - (5/5)

			## ## ## ## ## ## ## ## ## ## ## ## ##	9 11 11 11 11 11 11 11 11 11 11 11 11 11	11 12 13 13 13 13 13 14 14 14	11 11 11 11 11 11 11 11 11 11 11 11 11	11	
SEQUENCE ID NUMBER	GENOTYPE							
H H H H H	## ## ## ## ## ## ## ## ## ## ## ## ##	13 13 14 14 17	61 01 01 11 11 11 11 11 11 11	11 11 11 11 11 11 11 11	11 11 11 11 11 11 11 11 11	\$1.19\$2.221.210.612.214.614.614.614.614.614.614.614.614.614.6	# 10 11 11 11 11 11	
1	19	281	GACTTAGICG	TTATCTGTGA	AAGCGCGGGG	GICCAGGAGG 1	ACGCGGCGAG	CCTGAGAGCC
~		281	GACTTAGICG	TTATCTGTGA	AAGTGCGGGG	GTCCAGGAGG 1	ACGCGGCGAG	CCTGAGAGCC
m		281	GACTTGGTCG	TTATCTGTGA	GAGTGCGGGG	GICCAGGAGG	ACGCGGCGAG	CCTGAGAGCC
đ		281	GACTTAGTCG	TTATCTGTGA	GAGTGCGGGA	GTCCAGGAGG	ACGCGGCGAA	CITGAGAGCC
S		281	GACTTAGTCG	TTATCTGTGA	AAGTCAGGGA	GTCCAGGAGG	ATGCAGCGAA	CCTGAGAGCC
9		281	GACCTAGTCG	•	TTATCTGCGA AAGTGCGGGG	GTCCAGGAGG	ACGCGGCGAG	CCTGAGAGCC
	######################################	281	GACCTTGTCG		erasserrarererenamentarerererererererererererererererererere		ammemmemmem ACGCGGCAAG	CCTACGAGCC
- ω		281	GACCTTGTCG		AAGCGCGGGA		ATGCGGCGAG	CCTACGAGTC
6		281	GACCTTGTCG	-	AAGCGCGGGA	ACCCAGGAGG	ACGCGGCGAA	CCTACGAGIC
10		281	GACCTTGTCG	TTATCTGCGA	GAGCGCGGGA	ACCCAAGAGG	ACGCGGCGAG	CCTACGAGIC
11		281	GACCTTGTCG	TTATCTGTGA	GAGCGCGGGA	ACCCAAGAGG	ACGCGGCGAG	CCIACGAGIC
12		281	GACCTTGTCG	•	TTATCIGIGA GAGCGCGGG	ACCCAAGAGG		CCIACGAGIC
11 11 11 11 11 11 11	li I	 7 6 6 6						11 11 11 11 11 11 11 11 11 11 11 11 11
L L	1110	107			TCAICICAGA AAGCCAGGGGGGGGGGGGGGGGGGGGGGGG		そのないのないかん	
უ- L -ქ ი		107			מייייייייייייייייייייייייייייייייייייי		**************************************	ひたりなりなりたりこ
13		197	DACCIOGITO		ICATUTAGA GAGICAGGG		ATOMOCOGAN.	いたのよりよりにして
16		281	GACCTAGICG		TCATCTCAGA GAGTCAAGGG		ATGAGCGAAA	CCTGAGAGCT
17			GACCTGGTCG	rcarcicega	TCATCTCGGA GAGCGAAGGT	AACGAGGAGG	ACGAGCGAAA	CCIGAGGCI
 		281	GATCTTGTGG	CCATTTGCGA	GAGCCAGGG	GATCTTGTGG CCATTTGCGA GAGCCAGGG ACGCACGAGG ATAAAGCGAG	ATAAAGCGAG	CCTGAGAGCC
19		281	ACCTIGGIGG		GAGCCAAGGG	CCATITICGA GAGCCAAGGG ACGCACGAGG	ATGAAGCGTG	CCTGAGAGTC
	# # # # # # # # # # # # # # # # # # #	186				neerensserensensensaarensensserensserensserensserensensensensensen Tegnestyska gastgatigs stoskogasig atagascasc CCTGAGAGG	ATAGAGCAGC	areconnancementare and areas and areas and areas and areas areas areas areas areas areas areas areas areas are
3	10	1 1	200000000000000000000000000000000000000					
2.1		281	GATCTGGTTG		GAGTGATGGC	GTCGACGAGG		
22		281	GATCTGGTTG		GAGTGATGGC	TEGTEGETEA GAGTEATEGE GTCAATGAGE	ATAGAGCAGC	CCTGGGAGCC
# # # # # # # # # # # # # # # # # # #	11 11 11 11 11 11 11 11 11 11	11 11	11	11 11 11 11 11 11 11 11 11	n 11 11 11 11 11 11 11 11		\$1 18 18 19 19 19 19 19	## ## ## ## ## ## ## ## ## ## ## ## ##

340 TOTAL

7/2/

Fig.

ENVELOPE REGION

	11 11 11 11 11 11 11 11 11 11	10 11 11 12 13 14 14 15 16 17 18 18 18 18 18 18 18 18 18 18 18 18 18			;		
i 11	ATATGGTGGC ACATGGTGGC ACATGGTGGC	ACATAATAGC ACATAATAGC ACGTGCTAGC	CTATG CTCCTGGCAT ACTTGGTGG CATCCCGGAG GTCATCCTGG ACATTATCAC	CTCAC TGGGGAGTCC TGGCGGCAT AGCGTATTTC CCCAC TGGGGAGTCC TGGCGGGCAT AGCGTATTTC CCCAC TGGGGAGTCC TGGCGGGCAT AGCGTATTTC			
TTGG ATGG GTGG	GIGGIAICGC AGITACICCG GAICCCACAA GCCGICAIGG ATAIGGIGGC GIGGIGICGC AGITACICCG GAICCCACAA AGCAICGIGG ACAIGGIGGC GIGGIGICGC AGITACICCG GAICCCGCAA GCIGICGIGG ACAIGGIGGC	GTATG GIGGIGGCGC ACGICCIGCG TITGCCCCAG ACCITGITCG ACATAAIAGC GIAIG GIGGIAGCAC ACGICCIGCG ICIGCCCCAG ACCIIGIICG ACAIAAIAGC GIAIG GIGGIGGCGC AAGICCIGCG IIIGCCCCAG ACCIIGIICG ACGIGCIAGC	CTAIG CTCCTGGCAT ACTTGGTGCG CATCCCGGAG GTCATCCTGG ACATTATCAC			5 U U	
CGTTG GTAATGGCTC AGCTGCTCCG GATCCCACAA GCCATC CGTTG GTGGTAGCTC AGGTACTCCG GATCCCACAA GCCATC CGCTG GTAGTAGCTC AGCTGCTCAG GGTCCCGCAA GCCATC	GATCCCACAA GATCCCACAA GATCCCGCAA	TTTGCCCCAG TCTGCCCCAG	CATCCCGGAG	r AGCGIATTIC	GGGGGCCCAC TGGGAGTCC TGGCGGGCCT TGCCTACTAT GGGGGCCCAC TGGGAGTCC TGGCGGGCCT TGCTTACTAT GGGGGCCCAC TGGGGAATCC TAGCGGGTCT TGCCTACTAT	CGGGGCCCAT TGGGGCATCT TGGCGGGCTT GGCCTATTAC CGGGGCCCAT TGGGCATCT TGGCGGGCCT AGCTATTAC CGGGGCCCAT TGGGGGCATCT TGGCGGGCCT GGCCTATTAC	T GGCTTATTTC
AGCTGCTCCG AGGTACTCCG AGCTGCTCCG	AGTTACTCCG AGTTACTCCG AGTTACTCCG	GIGGIGGCGC ACGICCIGCG GIGGIAGCAC ACGICCIGCG GIGGIGGCGC AAGICCIGCG	ACTTGGTGC	TGGTGCTCAC TGGGGAGTCC TGGCGGGCAT TGGAGCCCAC TGGGGAGTCC TGGCGGGCAT TGGAGCCCAC TGGGGAGTCC TGGCGGGCAT	GGGGGCCCAC TGGGAAGTCC TGGCGGGCCT GGGGGCCCAC TGGGGAATCC TGGCGGGCCT GGGGGCCCAC TGGGGAATCC TAGCGGGTCT	CGGGGCCCAT TGGGGCATCT TGGCGGGCTT CGGGGCCCAT TGGGGCATCT TGGCAGGCCT CGGGGCCCAT TGGGGGCTCT TGGCGGGCCT	GGACAC TGGGGCGTGA TGTTTGGCCT
GTAATGGCTC GTGGTAGCTC GTAGTAGCTC	GACAGCCCTA GTGGTATCGC AGCAGCCCTA GTGGTGTCGC GGCAGCCCTA GTGGTGTCGC	GTGGTGGCGC GTGGTAGCAC GTGGTGGCGC	CICCIGGCA	TGGGGAGTCC	TGGGGAGTCC TGGGGAGTCC TGGGGAATCC	GCCCAT TGGGGCATCT GCCCAT TGGGGCATCT GCCCAT TGGGGCATCT	C TGGGGCGTG
GACGCCGTTC GACGCCGTTC AACGCCCTCG	GACAG	TGTGG	TACCACTATO	TGGTG TGGAG	1) 	8) It	GGGA
		; ; ; ; ;		61 61 61	61 61 61 61	61 61 61	
	611	QIV	GII	# I D # I	0 I I I I I I I I I I I I I I I I I I I	GIV	
23 24 25	26 27 28	30	32	i <u>ii</u>	26 27 28 28	29 30 31	32

100 Total

Fig. 4a

SEQUENCE							
				01 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	77 11 11 11 11 11 11 11 11 11 11 11 11 1		11 14 15 15 15 15 15 15 15 15 15 15 15 15 15
	ID GI		GITAGIATGA GIGICGIGCA GCC	GCCTCCAGGA	ככככככבוככ	CGGGAGAGCC ATAG	ATAGTGGTCT
34		-	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA	CTCCAGGA	CCCCCCCICC		ATAGTGGTCT
35		 1	GTTAGTATGA GTGTCGTGCA GCC	GCCTCCAGGA			ATAGTGGTCT
36		~	GTTAGTATGA GTGTCGTGCA GCC	GCCTCCAGGA	כככככככבככ		ATAGTGGTCT
37		-	GTTAGTATGA GTGTCGTGCA GCC	GCCTCCAGGA			ATAGTGGTCT
38		-	GTTAGIAIGA GIGICGIGCA GCCICCAGGA	CTCCAGGA	ככככככבבככ	CGGGAGAGCC ATAC	ATAGTGGTCT
39	SECTERATE 1	13 11 11	GITAGTATGA GIGICGIGCA	GCCTCCAGGA		GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT	ATAGIGGICI
40	•		GTTAGTATGA GTGTCGTGCA GC	GCCTCCAGGA	CCCCCCTCC CGGGAGAGCC		ATAGTGGTCT
41		7	GTTAGTATGA GTGTCGTGCA GC	GCCTCCAGGA			ATAGTGGTCT
42		-	GTTAGTATGA GTGTCGTGCA GC	GCCTCCAGGA	CCCCCCCTCC		ATAGTGGTCT
43		-		GCCTCCAGGA	CCCCCCTCC		ATAGTGGTCT
44		-	GTTAGTATGA GTGTCGTGCA GC	GCCTCCAGGA	CCCCCCLCC	CGGGAGAGCC ATA	ATAGTGGTCT
45		-	GITAGIAIGA GIGICGIGCA GCCICCAGGA	CTCCAGGA	CCCCCCICC	CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT	STGGTCT
46	19 19	accessors II	GCTAGTATCA GTGTCGTACA GCCTCCAGGC CCCCCCTCC CGGGAGAGCC ATAGTGGTCT	CTCCAGGC	CCCCCCTCC	CGGGAGAGCC ATA	Gracies
47		-	GITAGIAIGA GICICGIACA GCCICCAGGC CCCCCCTCC CGGGAGAGCC AIAGIGGICT	CTCCAGGC	ວວນວວວວວວ	CGGGAGAGCC ATA	GTGGTCT
48		" # # #	IV 1 GITAGIACGA GIGICGIGCA GCCICCAGGA CICCCCCIC CGGGAGAGCC ATAGIGGIC	CTCCAGGA	CTCCCCCTCC	CGGGAGAGCC ATA	GTGGTCT
49		-	GITAGIACGA GIGICGIGCA GCCICCAGGA CCCCCCTCC	CTCCAGGA	ככככככבכב	CGGGAGAGCC ATA	ATAGIGGICI
500	eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee	;;	nander i GITAGTATGA GIGICGAACA GCCICCAGGA CCCCCCICC CGGGAGAGCC ATAGIGGICI V	CTCCAGGA		CGGGAGAGCC ATA	GIGGICT
51		~	GITAGIAIGA GIGICGAACA GCCICCAGGA CCCCCCCICC CGGGAGAGCC ATAGIGGICI	CCTCCAGGA	CCCCCCCTCC	CGGGAGAGCC ATA	GIGGICT

9/2/

Fig. 4t

5'UT Region (2/5)

SEQUENCE				
ID NUMBER	GENOTYPE	11 11 11 11	# # # # # # # # # # # # # # # # # # #	11 11 11 11 11 11 11 11 11 11 11 11 11
33	61	1	1	A GGACGACCGG GTCCTTTCTT GGATCAACCC
) E	;	61		GGACGACCGG GTCCTTTCTT
96	•	61	GCGGAACCGG TGAGTACACC GGAATTGCCA	GGACGACCGG GICCTITCII
9 6		61	GCGGAACCGG TGAGTACACC GGAATTGCCA	GGACGACCGG GTCCTTTCTT
		61	TGAGTACACC	GGACGACCGG GTCCTTTCTT
38		61	TGAGTACACC	GGACGACCGG GTCCTTTCTT
	nannanna GIT	61	GCGGAACCGG TGAGTACACC GGAATTGCCA	GGACGACCGG GICCITICIT
. 4	•	61	GCGGAACCGG TGAGTACACC GGAATTGCCA	GGACGACCGG GICCTITCII
7		61		CA GGACGACCGG GICCITICII GGAICAACCC
4.1		61		GGACGACCGG GICCITICIT
4.2		19		GGACGACCGG GTCCTTTCTT
4		61	GCGGAACCGG TGAGTACACC GGAATTGCCA	GGACGACCGG GTCCTTTCTT
45		61	GCGGAACCGG TGAGTACACC GGAATTGCCA	CA GGACGACCGG GICCTITCIT GGATCAACCC
11 13 13 10 11 15 11	11 15 19 11 11 11 11 11	. 11 11 11	13 13	######################################
46	GIII	61	GCGGAACCGG IGAGIACACC GGAATIGCCG GGAAGACIGG	GICTITCII
47		61	GCGGAACCGG TGAGTACACC GGAATTGC	TGAGTACACC GGAATTGCTG GGAAGACTGG GTCCTTTCTT GGATAACCC
		===	BERRERERES BER	GGGTGACCGG GTCCTTTCTT
. 4. 0	; }	61	GCGGAACCGG TGAGTACACC GGAATCGC	GGAATCGCTG GGGTGACCGG GTCCTTTCTT GGAGTAACCC
# # # # # # # # #	() () () ()	11 11 :	11 13 18	THE TARGET TO THE TENTH OF THE TRANSPORT
20	gv GV	0.1	GCGGAACCGG IGAGIACACA GGAALLACACAGAALLACACAGAALACACAGAALACACAGAALACACAGAAAAAAAA	
ב		61	GCGGAACCGG TGAGIACACC GGAALIGCCG GGAIGACCGG	*******

SUBSTITUTE SHEET

Fig. 4

5'UT Region (3/5)

	H H H H H H H					11 11 11 11 11 11 11 11 11 11 11 11 11		## ## ## ## ## ## ## ## ## ## ## ## ##
CGCGAGACTG CTAGCCGAGT AGTGTTGGGT	CTAGCC	CGCGAGACTO	GCICAAIGCC CGGAGAITIG GGCGIGCCCC	CGGAGATTT		121		51
GV 121 GCTCAATGCC CGGAGATTIG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGT	CTAGCC	CGCGAGACTG	cecerecec	CGGAGATTTG	GCTCAATGC	121		200
GCTCAATACC CAGAAATTIG GGCGIGCCCC CGCGAGAICA CTAGCCGAGI AGIGIIGGGI	CTAGCC	CGCGAGATCA	CCCTCCCC	CAGAAATTTG	GCTCAATAC	121		49
GCTCAATACC CAGAAATTIG GGCGTGCCCC CGCGAGATCA CTAGCCGAGT AGTGTTGGGT	CTAGCC	CGCGAGATCA	occereccc	CAGAAATTTG	1	121	GIV	48
ACTCTATGCC CAGCCATTTG GGCGTGCCC CGCAAGACTG CTAGCCGAGT AGCGTTGGGT	CTAGCC	CGCAAGACTG		CAGCCATTIG		121		47
ACTOTATIGC COGCCATTIG GOOGTGCCCC CGCAAGACTG CTAGCCGAGT AGCGTTGGGT	CTAGCC	CGCAAGACTG	GGCGTGCCCC	CGGCCATTIG		1	GIII	46
GCTCAATGCC TGGAGAITTG GGCGTGCCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT	CTAGCCGAGT	CGCGAGACTG	GCTCAATGCC TGGAGATTTG GGCGTGCCCC	TGGAGATTTG	GCTCAATGCC	121	; ; ; ; ; ;	45
BAGT AGTGTTGGGT	CIAGCCGAGT	CGCGAGACTG	TGGAGATITIG GGCGTGCCCC		GCTCAATGCC	121		44
SAGT AGTGTTGGGT	CTAGCCGAGT	CGCGAGACTG	Gecerecec	TGGAGATTTG	GCTCAATGCC	121		43
SAGT AGTGTTGGGT			CGCGTGCCCC	TGGAGATTTG	GCTCAATGCC	121		42
SAGT AGTGTTGGGT		CGCGAGACTG	GCTCAATGCC TGGAGATTTG GGCGTGCCCC	TGGAGATTTG	GCTCAATGCC	121		41
SAGT AGTGTTGGGT	CTAGCCGAGT	CGCGAGACTG	TGGAGATITG GCCGTGCCCC	TGGAGATTTTG	GCTCAATGCC	121		40
GCTCAATGCC TGGAGATTTG GGCGTGCCC CGCGAGACTG CTAGCCGAGT AGTGTTGGGT	CTAGCCG	CGCGAGACTG	GCTCAATGCC TGGAGATTG GGCGTGCCCC CGCGAGACTG	TGGAGATTTG	GCTCAATGCC	121		
AGT AGTGTTGGGT	CIAGCCGAGI	CGCAAGACTG	GGCGTGCCC	TGGAGATTTG	GCTCAATGCC	121		38
AGT AGTGTTGGGT	CTAGCCGAGT	CGCAAGACTG	CCCTCCCC	GCTCAATGCC TGGAGATTTG	GCTCAATGCC	121		37
AGT ACTGTTGGGT	CTAGCCGAGT	CGCGAGACTG	CCCTCCCC	GCTCAATGCC TGGAGATTTG	GCTCAATGCC	121		36
AGT ACTGTTGGGT	CTAGCCGAGT	CGCAAGATCA	GGCACGCCCC	GCTCAATGCC TGGAGATTTG	GCTCAATGCC	121		35
AGT ACTGTTGGGT	CTAGCCGAGT	CGCAAGACTG	CCCTCCCC	GCTCAATGCC TGGAGATTTG	GCTCAATGCC	121		34
AGT AGTGTTGGGT	CTAGCCGAGT	CGCAAGACTG	GGCGTGCCCC	GCTCAATGCC TGGAGATTTG	GCTCAATGCC	ı	! ! !	33
	; ; ; ;	1 1 1 1 1 1 1 1 1	1 1 1 1 1 1			1	GENOTYPE	ID NUMBER
								SEQUENCE
							13 14 11	

Fig. 4d

ENVELOPE REGION (4/5)

): (i	" Ненененененененененененененененене		13 CH III	GT
() () () () () () () () () () () () () (GAGGTCTCGT GAGGTCTCGT GAGGTCTCGT GAGGTCTCGT GAGGTCTCGT	GAGGTCTCGT GAGGTCTCGT GAGGTCTCGT GAGGTCTCGT GAGGTCTCGT GAGGTCTCGT	GAGGTCTCGT GAGGTCTCGT	GAGGTCTC
13 13 13 13 13 14 14 14 14 14 14 14	AGTGCCCCGG AGTGCCCCGG AGTGCCCCGG AGTGCCCCGG AGTGCCCCGG	AGTGCCCCGG AGTGCCCCGG AGTGCCCCGG AGTGCCCCGG AGTGCCCCGG AGTGCCCCGG	AGTGCCCCGG AGTGCCCCGG	AGTGCCCCGG AGTGCCCCGG
11 11 11 14 11 11 11 11	GGTGCTTGCG 1 GGTGCTTGCG 1 GGTGCTTGCG 1 GGTGCTTGCG 1 GGTGCTTGCG 1	TGCCTGATAG GGTGCTTGCG AGTGCCCCGG	GGTGCTTGCG	GGTGCTTGCG
TYPS THE THE THE THE THE THE THE THE THE THE	GAAAGGC CITGIGGTAC IGCCIGATAG GGIGCITGCG AGIGCCCCGG GAGGICICGT GAAAGGC CITGIGGGTAC IGCCIGATAG GGIGCITGCG AGIGCCCCGG GAGGICICGT	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCCGG GAGGTCTCGT CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT CGCGAAAGGC CTTGTGGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT CGCGAAAGGC CTTGTGGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCGG GAGGTCTCGT CGCGAAAGGC CTTGTGGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	181 TGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT CGCGAAAGGC CTTGTGGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT
11 21 41 42 8 8 9	CTTGTGGTAC CTTGTGGTAC CTTGTGGTAC CTTGTGGTAC CTTGTGGTAC	CTTGTGGTAC CTTGTGGTAC CTTGTGGTAC CTTGTGGTAC CTTGTGGTAC CTTGTGGTAC	CTTGTGGTAC	CTTGTGGTAC
	CGCGAAAGGC CGCGAAAGGC CGCGAAAGGC CGCGAAAGGC CGCGAAAGGC	CGCGAAAGGC CGCGAAAGGC CGCGAAAGGC CGCGAAAGGC CGCGAAAGGC	TGCGAAAGGC TGCGAAAGGC	CGCGAAAGGC
; ; ; ;	181 181 181 181 181	181 181 181 181 181 181	181	181
GENOTYPE	91	IID	6111	OID
QUENCE NUMBER		39 44 42 44 45 5	46	48

12/2/

Fig. 4e

5'UT Region (5/5)

252 Total

Fig. 58

CORE REGION

SEQUENCE
ID NUMBER GENOTYPE

52	19		ATGAGCACGA	ATCCTAAACC	ATGAGCACGA ATCCTAAACC TCAAAAAAA AACAAGGTA ACACCAACGG TCGCCCACAG	AACAAACGTA	ACACCAACCG	1 ATGAGCACGA ATCCTAAAAC TCAAAAAAA AACAAACGTA ACACCAACCG TCGCCCACAG
53		-	ATGAGCACGA	ATCCTAAACC	ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAAACG TCGCCCACAG	ACCAAACGTA	ACACCAACCG	TCGCCCACAG
54		-	ATGAGCACGA	ATCCTAAACC	ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAAACG TCGCCCACAG	ACCAAACGTA	ACACCAACCG	TCGCCCACAG
55			ATGAGCACGA	ATCCTAAACC	ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG TCGCCCACAG	ACCAAACGTA	ACACCAACCG	TCGCCCACAG
56		~	ATGAGCACGA	ATCCTAAACC	ATGAGCACGA ATCCTAAACC TCAAAGAAGA ACCAAACGTA ACACCAAACC TCGCCCACAG	ACCAAACGTA	ACACCAACCG	TCGCCCACAG
57		-	ATGAGCACGA	ATCCTAAACC	ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG TCGCCCACAG	ACCAAACGTA	ACACCAACCG	TCGCCCACAG
	IID	11 27 27 28 41 41 41 41 41 41 41 41 41 41 41 41 41	PSSEREGEBERGESSESSESSESSESSESSESSESSESSESSESSESSESS	ATCCTAAACC	o=====================================	ACCAAACGIA	ACACCAACCG	CCGCCCACAG
59		-	ATGAGCACAA	ATCCTAAACC	ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG	ACCAAACGTA	ACACCAACCG	CCGCCCACAG
9	•	-	ATGAGCACAA		ATCCTAAACC CCAAAGAAAA ACCAAACGTA ACACCAACCG	ACCAAACGTA	ACACCAACCG	TCGCCCACAG
61		-	ATGAGCACGA		ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG	ACCAAACGTA	ACACCAACCG	CCCCCACAG
62			ATGAGCACGA		ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG	ACCAAACGTA	ACACCAACCG	CCGCCCACAG
. 63		-	ATGAGCACGA		ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG	ACCAAACGTA	ACACCAACCG	CCGCCCACAG
64		-	ATGAGCACGA	ATCCTAAACC	ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG CCGCCCACAG	ACCAAACGTA	ACACCAACCG	CCGCCCACAG
# # # # # # # # # # # # # # # # # # #		# # # # # # # # # # # # # # # # # # #						
a a	6111	-4 -	ATGAGCACAA	ATCLTAAACC	TCAAAGAAAA	ACCARAGAR	ACACTAACCO	AIGAGCACAA AICCIAAAAC ICAAAGAAAA ACCAACIAACCG CCGCCCACAG
99			ATGAGCACAA	ATCCTCAACC	TCAAAGAAAA	ACCAAAAGAA	ACACTAACCG	atgagcacaa atcctcaacc tcaaagaaaa accaaaagaa acactaaccg ccgcccacag

Fig.

SEQUENCE ID NUMBER GENOTYE	GENOTYPE	# N U U		1) 1) 1) 1) 1) 1) 1) 1) 1) 1)	# # # # # # # # # # # # # # # # # # #			
52 GI 53 54 55 56 57		61 61 61 61 61	g #1	TCCCGGGTGG TCCCGGGTGG TCCCGGGTGG TCCCGGGTGG TCCCGGGTGG	TCCCGGGTGG CGGTCAGATC TCCCGGGTGG CGGTCAGATC TCCCGGGTGG CGGTCAGATC TCCCGGGTGG CGGTCAGATC TCCCGGGTGG CGGTCAGATC	GTTGGTGGAG GTTGGTGGAG GTTGGTGGAG GTTGGTGGAG	TTTACTEGET TTTACTEGET TTTACTEGET TTTACTEGET TTTACTEGET TTTACTEGET	GACGTCAAGT TCCCGGGTGG CGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG GACGTCAAGT TCCCGGGTGG CGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG GACGTTAAGT TCCCGGGTGG CGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG GACGTCAAGT TCCCGGGTGG CGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCCCAGG
5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	011	1	4	10000000000000000000000000000000000000	TCCCGGGCGG TGGCCAGGTC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG TCCCGGGCGG TGGTCAGATC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG TCCCGGGCGG TGGTCAGATC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG TCCCGGGCGG TGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG TCCCGGGCGG TGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG TCCCGGGCGG TGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG	GTTGGTGGAG GTTGGTGGAG GTTGGTGGAG GTTGGTGGAG GTTGGTGGAG	TTTACCTGTT TTTACCTGTT TTTACCTGTT TTTACTTGTT TTTACTTGTT TTTACCTGTT TTTACCTGTT	GACGTLAAGT TCCCGGGCGG TGGTCAGGTC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG GACGTCAAGT TCCCGGGCGG TGGTCAGATC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG GACGTCAAGT TCCCGGGCGG TGGTCAGATC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG GACGTCAAGT TCCCGGGCGG TGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG
99	GIII	I 61	1	TCCCGGGCGG	GACGICAAGI ICCCGGGCGG IGGCCAGAIC GIIGGCGGAG IAIACIIGCI GCCGCGCAGG GACGICAAGI ICCCGGGCGG IGGICAGAIC GIIGGCGGAG IAIACIIGII GCCGCGCAGG	GTTGGCGGAG GTTGGCGGAG	TATACTIGCT TATACTIGIT	GCCGCGCAGG

Fig. 5c

CORE REGION (3/9)

GENOTYPE

SEQUENCE ID NUMBER

25	GI	121	GCCCTAGAT	TGGGTGTGCG	GGCCCTAGAT TGGGTGTGCG CGCGACGAGA AAGACTTCCG AGCGGTCGCA ACCTCGAGGT	AAGACTTCCG	AGCGGTCGCA	ACCTCGAGGT
53		121	GGCCCTAGAT	TGGGTGTGCG	GGCCCTAGAT TGGGTGTGCG CGCGACGAGG AAGACTTCCG AGCGGTCGCA ACCTCGAGGT	AAGACTTCCG	AGCGGTCGCA	ACCTCGAGGT
54		121	GCCCTAGAT		TGGGTGTGCG CGCGACGAGG AAGACTTCCG AGCGGTCGCA ACCTCGAGGT	AAGACTTCCG	AGCGGTCGCA	ACCTCGAGGT
52		121	GGCCCTAGAT	TGGGTGTGCG	CACGACGAGG	AAGACTTCCG	CACGACGAGG AAGACITCCG AGCGGICGCA ACCICGAGGI	ACCTCGAGGT
56		121	GGCCCTAGAT	TGGGTGTGCG	CGCGACGAGG	AAGACTTCCG	CGCGACGAGG AAGACTTCCG AGCGGTCGCA ACCTCGAGGT	ACCTCGAGGT
21		121	GGCCCTAGAT	Tecetetece	GGCCCTAGAT TGGGTGTGCG CGCGACGAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGT	AAGACTTCCG	AGCGGTCGCA	ACCTCGTGGT
11 11 11 11 11 11	11 11 11 11 11 11 11 11 11 11 11 11 11	11 11 11 11 11 11 11 11 11 11 11 11 11					11 11 11 11 11 11 11 11 11 11 11 11 11	
58	611	121	GGCCCCAGGT	TGGGTGTGCG	GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA	AAGACTTCCG	AGCGGTCGCA	ACCTCGTGGA
59		121	GGCCCCAGGT	TGGGTGTGCG	GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA	AAGACTTCCG	AGCGGTCGCA	ACCTCGTGGA
09	•	121	GGCCCCAGGT	GCCCCAGGT TGGGTGTGCG		AAGACTTCCG	CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA	ACCTCGTGGA
61		121	GGCCCCAGGT	Géceceager recereree		AAGACTTCCG	CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA	ACCTCGTGGA
.62		121	GGCCCCAGGT	recererece		AAGACTTCCG	CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA	ACCTCGTGGA
63		121	GCCCCAGGT	recererece	CGCGACTAGG	AAGACTTCCG	CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA	ACCTCGTGGA
64		121	GGCCCCAGGT	TGGGTGTGCG	TEGETETEC CECEACTAGE AAGACTICCE AGCEGICGEA ACCICGIGGA	AAGACTTCCG	AGCGGTCGCA	ACCTCGTGGA
12 17 12 12	11 18 11 11 11 11 11	U H	## ## ## ## ## ## ## ## ##			## ## ## ## ## ## ## ## ## ## ## ## ##	11 11 11 11 11 11 11 11	11 11 11 11 11 11 11 11 11
65	GIII	121	GGCCCGAGAT	recererece	GGCCCGAGAT TGGGTGTGCG CGCGACGAGG AAAACTTCCG AACGATCCCA GCCACGCGGA	AAAACTTCCG	AACGATCCCA	GCCACGCGGA
99		121	GGCCCCAGGT	TGGGTGTGCG	GGCCCCAGGT TGGGTGTGCG CGCGACGAGG AAAACTTCCG AACGGTCCCA GCCACGTGGG	AAAACTTCCG	AACGGTCCCA	GCCACGIGGG

SUBSTITUTE SHEET

GENOTYPE

ID NUMBER SEQUENCE

CORE REGION (4/9)

AGGCGCCAGC CCATCCCCAA AGATCGGCGC ACCACTGGCA AGTCCTGGGG GAAGCCAGGA TCAGCCCGGG CCATCCCTAA AGATCGTCGC ACCGCTGGCA AGTCCTGGGG AAGGCCAGGA TCAGCCCGGG GGACCIGGGC ICAGCCCGGG CGGCCTGGGC CGGCCTGGGC GGACCTGGGC GGACCTGGGC GGACCTGGGC GGTCCTGGGC GGCCTGGGC GGGCCTGGGC GGGCCTGGGC CGGCCTGGGC GGACCTGGGC GGACCTGGGC AGGCGACAAC CTATCCCCAA GGCTCGCCAG CCCGAGGGCA CCCGAGGGCA AGGCGACAAC CIATCCCCAA GGCTCGCCAG CCCGAGGGCA CCCGAGGCCA CCCGAGGGCA CCCGAGGGTA CCCGAGGGCA CCCGAGGGCA CCCGAGGGCA CCCGAGGGCA CCCGAGGGCA CCCGAGGGTA CCCGAGGGCA GGCTCGCCGG GGCTCGCCAG GCTCGCCGG SGCTCGCCGG GGCTCGTCGG Secentres GCCCGTCGG GCTCGTCGA GGCACGTCGG GGGGGGTCGG GGCTCGCCAG CTATCCCCAA CTATCCCCAA CTATCCCCAA CTATCCCCAA CTATCCCCAA CTATCCCCAA CCATCCCCAA CTATCCCCAA CTATCCCCAA CTATCCCTAA CTATCCCCAA AGGCGACAAC AGGCGTCAGC AGGCGACAAC AGGCGACAAC AGGCGACAAC AGGCGACAAC AGACGICAGC AGACGTCAGC AGACGCCAGC AGACGICAGC AGACGTCAGC AGACGTCAGC 181 181 181 181 181 181 181 181 181 181 181 181 181 61 61 61 61 61 61 61 61 61 61 61 61 54 55 99 57 59 9 61 29

SUBSTITUTE SHEET

Fig. 5e

SEQUENCE ID NUMBER GENOTY	11 Ca 1	17 1 14 1 14 1 15 1 16 1			0 0 0 0 0 0 0 0 0 0	9 11 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		17 16 19 19 19 19 19 19
25.	19	241	TACCCTTGGC CC	CTCTATGG	CAATGAGGGC	receereed	TACCCTTGGC CCCTCTATGG CAATGAGGGC TGCGGGTGGG CGGGATGGCT CCTGTCTCCC	TGTCTCCC
55 S		241 241	TACCCTTGGC CC	CCCTCTATGG	CAATGAGGGT TAATGAGGGT	CAATGAGGGT TGCGGGTGGG TAATGAGGGT TGCGGATGGG	CCCTCTATGG CAATGAGGGT TGCGGGTGGG CGGGATGGCT CCTGTCTCCC CCCTCTATGG TAATGAGGGT TGCGGATGGG CGGGATGGCT CCTGTCCCCC	ccrercrecc
55		241	TACCCTTGGC CC	CTCTATGG	CCCTCTATGG CAATGAGGGC TGCGGGTGGG		CGGGATGGCT CC	CCTGTCTCCC
56		241	TACCCTIGGC CC	CTCTATGG	CCCTCTATGG CAATGAGGGT TGCGGGTGGG		CGGGATGGCT	CCTGTCTCC
57		241	_	CTCTATGG	CCCICIATGG CAATGAGGGT TGCGGGTGGG CGGGATGGCT	receestese	CGGGATGGCT CC	
# 25 # 50 # 50 # 50 # 50 # 60 # 60 # 60 # 60 # 60 # 60 # 60 # 6	IID	241	81 13	CTCTATGG	CAATGAGGGT	ATGGGGTGGG	TACCCTTGC CCCTCTATG CAATGAGGT ATGGGGTGG CAGGATGGCT CCTGTCACC	TGTCACCC
59	•	241	TACCCTTGGC CC	CCCTCTATGG	CAACGAGGGT	ATGGGGTGGG	CAGGATGGCT	CCTGTCACCC
09	•	241	TACCTTGGC CC	CCCTCTATGG	CAACGAGGGT	ATGGGGTGGG	CAGGATGGCT	CCTGTCACCC
61,		241	TACCCTTGGC CC	CCCTCTATGG	CAATGAGGGT	ATGGGGTGGG	CAGGGTGGCT	ccrercccc
62		241	TATCCTTGGC CC	CCCTCTATGG	CAATGAGGGT		CTGGGGTGGG CAGGATGGCT CC	CCTGTCACCC
63		241	TACCCTTGGC CC	CCCTCTATGG	CAATGAGGGT		ATGGGGTGGG CAGGATGGCT CC	CCTGTCACCC
64		241	TACCCCTGGC CCCTCTATGG CAATGAGGGT ATGGGGTGGG CAGGATGGCT	CTCTATGG	CAATGAGGGT	ATGGGGTGGG	CAGGATGGCT CC	CCTGTCACCC
		241	nonning and national	CTGTATGG	GAATGAGGGT	CTCGGCTGGG	contact contac	TGTCCCCC
99		241	TACCCTTGGC CC	CTGTATGG	GAATGAGGGT	ジンプルノンプノルノ	プランシン・サンプ トラング・アラン かっかん かっかん かんしゅう かかん かんしょうしょ かんしょうかん かっかん かんしゅう しゅうしゅう しゅう 	グランプラー

SUBSTITUTE SHEET

18/2/

Fig. 5

CORE REGION (6/9)

starraters solden serters of the state of the state of the series of the

ı	1						ti H							ti Ii		1
3G	CAATTIGGGT	CAATTTGGGT	CAATTTGGGT	CAATTIGGGT	CAATTIGGGT	CGIGGCICIC GGCCIAGCIG GGGCCCCACA GACCCCCGGC GIAGGICGCG CAAITIGGGI	RESERVED TO THE THE TREE TO THE TREE TO THE TREE TO THE TREE TO THE TREE TRAINED TRAINED TO THE TREE TRAINED TREE TRAINED TRAI	CGIGGCICIC GGCCIAGIIG GGGCCCCACG GACCCCCGGC GIAGGICGCG IAAIIIGGGI	GTAGGTCGCG TAATTTGGGT	GTAGGTCGCG TAATTTGGGT	GTAGGTCGCG CAACTTGGGT	GTAGGTCGCG CAATTTGGGT	GTAGGTCGCG TAATTTGGGT	TOTICETHER CONTRACTOR OF THE STATE OF THE ST	CARCITOGGI	301 CGCGGTTCTC GCCCTTCATG GGGCCCCACT GACCCCGGC ATAGATCACG CAACTTGGGT
1 1 1 1 1	GTAGGTCGCG	GTAGGTCGCG	GTAGGTCGCG	GTAGGTCGCG	GTAGGTCGCG	GTAGGTCGCG	GTAGGTCGCG	GTAGGTCGCG	GTAGGTCGCG			GTAGGTCGCG	GTAGGTCGCG		9797149414	ATAGATCACG
1 1 2 1 1 1 1 1	GACCCCCGGC	GACCCCCGGC	GACCCCGGC	GACCCCGGC	GACCCCCGGC	GACCCCGGC	GACCCCCGGC	GACCCCCGGC	GACCCCCGGC	GACCCCGGC	GACCCCGGC	GACCCCGGC	GACCCCCGGC		7007777780	GACCCCCGGC
9 9 1 1 1 1 1	CGTGGCTCTC GGCCTAGCTG GGGCCCCACA GACCCCCGGC GTAGGTCGCG CAATTTGGGT	CGTGGCTCTC GGCCTAGTTG GGGCCCCACA GACCCCGGC GTAGGTCGCG	GGCCTAGITG GGGCCCTACA GACCCCGGC	GGCCTAGCTG GGGCCCCACA GACCCCGGC	GGCCTAACTG GGGCCCCACA GACCCCCGGC GTAGGTCGCG	GGGCCCCACA	GGGCCCCACG	GGGCCCCACG	ceceectec eectaette eeececace eacceeec	cécecrece eceraerre eececeaca: eaceceeec	CGCGGCTCTC GGCCTAGCTG GGGCCCTACC GACCCCGGC	CGIGGIICIC GGCCIAGIIG GGGCCCCACG GACCCCGGC	CGCGGCTCCC GGCCTAGTIG GGGCCCCAAA GACCCCGGC	***************************************	てつなつつつうちち	GGCCCCACT
3 0 2 5 1 1 1 1	GCCTAGCTG	GGCCTAGITG	GGCCTAGTTG	GGCCTAGCTG	GGCCTAACTG	GGCCTAGCTG	GGCCTAGTTG	GCCCTAGTTG	GGCCTAGTTG	GGCCTAGITG	GGCCTAGCTG	GGCCTAGTTG	GGCCTAGTTG		014711777	GCCCTTCATG
	CGTGGCTCTC	CGTGGCTCTC	CGTGGCTCTC	CGIGGCICIC	CGCGGCTCTC	cereecrere	CGTGGCTCTC	CGIGGCICIC	ລລາວອວອວ	céceecrece	CGCGGCTCTC	CGIGGIICIC	CGCGGCTCCC		717700107	CGCGGTTCTC
! ! ! !	301	301	301	301	301	301	301	301	301	301	301	301	301	11 11 11 11 11 11 11 11 11 11 11 11 11	107	301
GENOTYPE	ß						IID		,						1110	
ID NUMBER GENOTYPE	52	53	. 54	52	56	24	11 12 12 12 12 12 12 12 12 12 12 12 12 1	59	09	61	62	63	64	## ## ## ## ## ## ## ## ## ## ## ## ##	0	99

Fig. 5g

Į

CORE REGION (7/9)

GENOTYPE

ID NUMBER

361 AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA TGGGGTACAT ACCGCTCGTC 361 AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCACA TGGGGTACAT ACCGCTCGTC 361 AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCACA TGGGGTACAT TCCGCTCGTT 361 AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA TGGGGTACAT ACCGCTCGTC 361 AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA TGGGGTACAT ACCGCTCGTC 361 AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA TGGGGTACAT ACCGCTCGTC	ŗ	1 11
AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCACA AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCACA AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA	AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA AAGATCATCG ATACCCTCAC GTGCGGCTTC GCCGACCTCA	GCCGACCTCA
AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCACA AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCACA AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA	AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA AAGGTCATCG ATACCCTCAC GTGCGGCTTC GCCGACCTCA	GTGCGGTTTT
ATACCCTTAC ATACCCTTAC ATACCCTCAC ATACCCTTAC ATACCCTTAC	ATACCCTCAC ATACCCTCAC ATACCCTCAC ATACCCTCAC ATACCCTTAC ATACCCTCAC	ATACCCTAAC ATACCCTAAC
AAGGTCATCG AAGGTCATCG AAGGTCATCG AAGGTCATCG AAGGTCATCG	AAGGTCATCG AAGGTCATCG AAGGTCATCG AAGGTCATCG AAGGTCATCG AAGGTCATCG AAGGTCATCG	AAGGTCATCG AAGGTCATCG
361 361 361 361 361	361 361 361 361 361 361	361
19		GIII
52 53 54 55 56	58 59 61 62 63	

Fig. 5h

ţ

CORE REGION (8/9)

Z	ωı							
======================================		421	H .	TTGGAGGCGC	TGCCAGGGCC	CTGGCGCATG	GECECCOTC TIGGAGGCGC TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAAGAC	TCTGGAAGAC
53		421	CGCGCCCTC	TTGGAGGCGC	TGCCAGGGCT	CTGGCGCATG	receaseser ersecsears secresser	TCTGGAAGAC
54		421	GGCGCCCTC	Treesesses		CTGGCGCATG	reccaegec creeceare ecerceger	TCTGGAAGAC
55		421	CCCCTC	TTGGAGGCGC		TGCCAGAGCC CTGGCGCATG	GCGTCCGGGT	TCTGGAAGAC
56		421	CGCGCCCTC	TTGGAGGCGC		TGCCAGGGC CTGGCGCATG	GCGTCCGGGT	TCTGGAAGAC
57		421	OGCGCCCTC	TTGGAGGCGC	TGCCAGGGCC	CTGGCGCATG	TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAAGAC	TCTGGAAGAC
	11 11 11 11 11 11 11 11 11 11 11 11 11	10 10 10 10 10						
	GII	421	ລລລລລລລອອອ	TTAGGGGCGC	TGCCAGGGCC	TIGGCGCAIG	GGCGCCCCCC TTAGGGGCGC TGCCAGGGCC TTGGCGCATG GCGTCCGGGT TCTGGAGGAC	TCTGGAGGAC
59		421	ລລວລລລອລອອ	TAGGGGGCGC		TGCCAGGGCC CTGGCACATG	GTCTCCGGGT	TCTGGAGGAC
60	•	421	ວວວວວວວວ	TAGGGGGCGC	TGCCAGGGCC	TGCCAGGGCC CTGGCACATG	GTGTCCGGGT	TCTGGAGGAC
61		421	ဝဇင္ဇဇင္ဇင္ဇင္ဇင္	TAGGGGGCGC	TGCCAGGGCC	TGCCAGGGCC CTGGCGCATG	GCGTCCGGGT	TCTGGAGGAC
62		421	ລລລລລລລອ	TTAGGGGCGC		receasesc crescerts	GCGTCCGGGT	TCTGGAGGAC
63		421	ວວວວວວວວອ	TAGGGGGCGC	TGCCAGGGCC	CIGGCGCATG	GOCGCCCCCC TAGGGGGCGC TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAGGAC	TCTGGAGGAC
64		421	LOCOCOCO	TAGGGGGCGC	TGCCAGGGCC	CTGGCGCATG	TAGGGGGGG TGCCAGGGCC CTGGCGCATG GCGTCCGGGT	TCTGGAGGAC
	IIID	421	922222299	TIGGAGGCGT	TGCCAGAGCT	CTCGCCCACG	GGCGCCCCCG TTGGAGGCGT TGCCAGAGCT CTCGCCCACG GAGTGAGGGT TCTGGAGGAT	TCTGGAGGAT
· •		7.23	じしししししじまじむ	上されなどかなどした山	THE THE THE TOTO TO THE CONTROL OF THE TOTO TOTO TOTO TOTO TO THE TOTO T	CHARCOCO	CCCTCAGGGT	TCTGGAAGAC

Fig. 5i

CORE REGION (9/9)

ID NUMBER GENOTYPE

SEQUENCE

11 11 11 11 11 11		81 81 81 81 81		11 11 11 11 11 11 11 11	11 11 11 11 11 11 11 11 11	11 13 14 28 28 28 29 21 21 21	11 12 12 13 14 14 15 16 17 18	11 13 13 13 13 13 14 14 15 14	## ## ## ## ## ## ## ## ## ## ## ## ##
52	GI		GGCGTGAACT	AIGCAACAGG GAACCTICCI GGIIGCICII ICICIAICII CCIICIGGCC CIGCICICI	GAACCTTCCT	GGTTGCTCTT	TCTCTATCTT	CCTTCTGGCC	Crecreter
53		481	GGCGTGAACT	ATGCAACAGG GAACCTTCCT	GAACCITCCI	GGTTGCTCTT	TCTCTATCTT	ccrrcreecc crecrcrcr	CIGCICICI
54		481	GGCGTGAACT	ATGCAACAGG	GAATCTTCCT	GGTTGCTCTT	TCTCTATCTT	CCTTCTGGCC	CTICICICI
55		481	GGCGTGAACT	ATGCAACAGG	GAACCITCCC	GGTTGCTCTT	TCTCTATCTT	CCTTCTGGCC	CTGCTCTCT
56		481	GGCGTGAACT	ATGCAACAGG	GAACCTTCCT	GGTTGCTCTT	TCTCTATCTT	CCTTCTGGCC	CIGCICICI
57		481	GGCGTGAACT	ATGCAACAGG GAACCTTCCT	GAACCTTCCT	GGTTGCTCTT	TTTCTATTT	ccrrcagcc crgcrcrcr	CIGCICICI
11 01 11 11 11 11	# # # # # # # #	11 11 11 11 11		11 00 00 00 00 00 00 00 00 00 00 00 00 0		11 11 11 11 11 11 11 11 11 11 11 11 11	1) 1) 1) 1) 1) 1) 1) 1) 1) 1)		11 11 11 11 11 11 11
58		481	GGCGTGAACT	GGCGIGAACI ACGCAACAGG GAAICIGCCC GGIIGCICCI IIICIAICII CCICIIGGCI	GAATCTGCCC	GGTTGCTCCT	TTTCTATCTT	CCICITGGCI	CIGCIGICC
59		481	GGCGTGAACT	GGCGTGAACT ATGCAACAGG GAATTTGCCC GGTTGCTCTT TCTCTATCTT	GAATTTGCCC	GGTTGCTCTT	TCTCTATCTT	CCTCTTGGCT	CIGCIGICC
9		481	GGCGTGAACT	GGCGTGAACT ATGCAACAGG GAATTTGCCT	GAATTTGCCT	GGTTGCTCTT	TCTCTATCTT	CCTCTTGGCT	CIGCIGICC
61		481	GGCGTGAACT	ATGCAACAGG	GAATCTGCCC	GGTTGCTCTT	TCTCTATCTT	CCTCTTGGCT	TIGCIGICC
62	÷	481	GGCGTGAACT	ATGCAACAGG	GAATTTGCCC	GGTTGCTCTT	TCTCTATCTT	CCICITGGCI	TIGCICICC
63		481	GGCGTGAACT	ATGCAACAGG	GAATCTGCCC	GGTTGCTCCT	TTTCTATCTT	CCITCIGGCI	TIGCIGICC
64		481	GGCGTGAACT	GGCGTGAACT ATGCAACAGG	GAATCTACCC GGTTGCTCTT	GGTTGCTCTT	TCTCTATCTT	CCICITGGCI	TIGCIGICC
11 11 11 11 11 11	11 14. 14. 10 11 11	## ## ## ## ## ## ##		11 12 12 13 14 14 15 16 17 17 18 18 18 18 18 18 18 18 18 18 18 18 18				0 6 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	11 11 11 11 11 11 11 11 11 11 11 11 11
9	GIII	481	GGGGTAAATT	GGGGTAAATT ATGCAACAGG GAATTTGCCC GGTTGCTCTT TCTCTATCTT TCTCTTAGCC CTCTTGTCT	GAATTTGCCC	GGTTGCTCTT	TCTCTATCTT	TCICITAGCC	CTCTTGTCT
99		481	GGGATAAATT	GGGATAAATT ATGCAACAGG GAATCTGCCC	GAATCTGCCC				
15 16 18 18 18 19 11	11 11 12 13 15 19 18 11	# } } } } }	0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	#	11 11 11 11 11 11 11 11	11 11 11 11 11 11 11 11	" " " " " " " " " " " " " " " " " " "	11 11 11 11 11 11 11 11 11	11 16 17 15 16 16 17 18

549 Total